CUa-H05948-25-BP 9726

CENTRAL LIBRARY
8-11-
UNIVERSITY OF CALCUTTA
ALIPORE CAMPUS

THE JOURNAL OF PHILOSOPHY

25

VOLUME CV, NUMBER 1
JANUARY 2008

page
5 *Is Ignorance Bliss?* Joseph B Kadane, Mark Schervish, and Teddy Seidenfeld
37 *Relative Expectation Theory* Mark Colyvan
45 NOTES AND NEWS
47 MEMO TO AUTHORS
48 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Michael J Mooney, *Treasurer,* Liaquat Ahamed, *Other Trustees* Lee Bollinger, Leigh S Cauman, Barbara Gimbel, Kent Greenawalt, Lynn Nesbit, Daniel Shapiro

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721
You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2008

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

KP2T25

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices
POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2008 by the Journal of Philosophy, Inc ISSN 0022-362X

CENTRAL LIBRARY
UNIVERSITY OF CALCUTTA
ACRE CAMPUS

THE JOURNAL OF PHILOSOPHY

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J E Woodbridge and J McKeen Cattell In 1906, Wendell T Bush became associated with the JOURNAL as co-editor The title was changed in 1921, and the JOURNAL was incorporated in the State of New York under its present name in 1923 Since its founding, it has been published from the Department of Philosophy at Columbia University It is typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

THE JOURNAL OF
PHILOSOPHY

VOLUME CV, NUMBER 1
JANUARY–DECEMBER 2008

Edited by
Bernard Berofsky, Akeel Bilgrami
John Collins, Arthur C. Danto
Kent Greenawalt, Patricia Kitcher, Philip Kitcher
Isaac Levi, Wolfgang Mann, Frederick Neuhouser
Christopher Peacocke, Carol Rovane
Achille C. Varzi

Consulting Editors
David Albert, James T. Higginbotham, Robert May
Charles D. Parsons, Wilfried Sieg

Managing Editor
John Smylie

Published by The Journal of Philosophy, Inc.

STATEMENT OF COPYRIGHT POLICY

We would like to draw your attention to the protection copyright affords against multiple photocopying by libraries, instructors, and copy centers. The 1976 copyright law provides that photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner—"fair use" meaning single copying for personal use Accordingly, the JOURNAL authorizes personal or educational multiple copying only if a request is made in writing, and the fee of $ 05 per page/per copy is paid directly to us *Each copy must include a notice of copyright* We shall turn over to the authors any substantial payments (over $30 00, an administrative fee) obtained from all photocopying Inquiries about photocopying may be sent to the Managing Editor Please consult our website, www journalofphilosophy org, for more information about our copyright policies

© Copyright 2008 by the Journal of Philosophy, Inc

ISSN 0022-362X

POSTMASTER Second class postage paid at New York, NY, and other mailing offices Please send address changes to THE JOURNAL OF PHILOSOPHY, 709 Philosophy Hall, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 Fax (212) 932-3721

THE JOURNAL OF PHILOSOPHY
VOLUME CV, NO. 1, JANUARY 2008

IS IGNORANCE BLISS?*

> “ where ignorance is bliss, ’tis folly to be wise ”
>
> —Thomas Gray

We are surrounded by homilies supporting the idea that human progress is intimately tied to the increase of our knowledge Research is justified that way, as is our enormous investment in education at all levels, as is our support of public libraries of all sorts Generally, we are willing, as a society, to pay a lot to make information available to ourselves and others

It is easy to see that there may be some kinds of information that one might rationally pay not to have, if paying that price would prevent others from learning the information as well For example, if there were a cheap and effective means of killing all living beings, we would have no use for that knowledge ourselves, but would pay a high price to prevent others who might want to use it from gaining the knowledge of how to do such a thing This paper focuses on a simpler problem, whether there is information a rational person would pay not to have for that person’s private use

The paper is organized as follows Section i reviews the standard Bayesian theorem about the non-negative expected worth of cost-free sample information Section ii describes a more detailed Bayesian model that permits a more subtle understanding of what “cost-free” means in this context Section iii gives formal definitions of the value of information and cost-free information based on the formulation in section ii Section iv shows by examples that a Bayesian might pay not to receive cost-free information if the prior is improper or not countably additive Section v reviews three extensions of the

---

* Research partially supported by NSF grant DMS-0139911

0022-362X/08/0501/5–36

Bayesian idea to sets of probabilities, and shows that each requires or permits paying not to receive sample information Section VI concludes with a general discussion of whether refusing to pay to be shielded from information is a criterion that should be imposed on a reasonable decision theory, and, if so, what its consequences would be Finally, the Appendix discusses the sense of Bayesian updating used in this paper

I THE EXPECTED VALUE OF SAMPLE INFORMATION

There is an appealing intuitive argument supporting the idea that one should not pay not to receive cost-free information Informally stated, it goes as follows Suppose that one faces a decision at time t_0, to be made under uncertainty In the absence of an offer to acquire new information, one is prepared to make decision d at t_0 with expected utility u_d One might imagine acquiring new cost-free information x at a later time t_1, but then still making decision d regardless of x From the perspective of the initial choice point at time t_0, this is the same as choosing d, and has expected utility u_d But it might be that at time t_1, given x, there is a better decision than d to choose Then from the perspective of the initial choice point, at t_0, this more complicated plan has a higher expected utility than does the choice of d Thus, having the information x can not be harmful to one's expected utility, and it might be helpful From the perspective of the initial choice, at t_0, the expected worth of the new information x cannot be negative

This argument can be made rigorous as follows Let $U(d,\theta)$ be your utility function, which depends on both your decision d and on $\theta \in \Theta$, the unknown state of the world You have a distribution that jointly describes your probabilities for the data $x \in \chi$ and for θ, with joint density $p(x,\theta)$ Assume that there is no cost associated with learning x Without the data x, you would choose d to maximize

$$\int_X \int_\Theta U(d,\theta)p(x,\theta)\,d\theta\,dx \tag{1}$$

If you were to learn the data x, you would maximize your utility with respect to your conditional distribution $p(\theta \mid x)$, that is, maximize

$$\int_\Theta U(d,\theta)p(\theta \mid x)\,d\theta, \tag{2}$$

which has expectation, with respect to the unseen value of x,

$$\int_X \left[\max_d \int_\Theta U(d,\theta)p(\theta \mid x)\,d\theta\right] p(x)\,dx, \tag{3}$$

where $p(x)$ is the probability distribution of the data x The intuitive argument above suggests that (3) is no smaller than (1)

To show this, let d^* be a maximizer of (1) [The argument works just as well, if such a d^* does not exist, for d^* to be an ε-maximizer of (1)] Then for each x in (2),

$$\max_d \int_\Theta U(d,\theta)p(\theta|x)\,d\theta \geq \int_\Theta U(d^*,\theta)p(\theta|x)\,d\theta \tag{4}$$

Integrating both sides of (4) with respect to x, yields

$$\begin{aligned} \int_X \left[\max_d \int_\Theta U(d,\theta)p(\theta|x)\,d\theta\right] p(x)\,dx &\geq \int_X \int_\Theta U(d^*,\theta)p(\theta|x)\,d\theta p(x)\,dx \\ &= \int_X \int_\Theta U(d^*,\theta)p(\theta,x)\,d\theta dx \\ &= \max_d \int_X \int_\Theta U(d,\theta)p(\theta,x)\,d\theta dx, \end{aligned} \tag{5}$$

as claimed This result is a familiar one in Bayesian theory, with a long history[1]

II BAYESIAN MODELS INCORPORATING PAYING NOT TO LEARN

I hold a ticket to a mystery play, and take a taxi to the theater The taxi driver knows who-done-it and offers (threatens?) to inform me unless the tip is sufficiently large Does it make sense to pay to avoid learning this information?

It is clear that there are circumstances in which I might pay, and that the result rehearsed in section I seems to contradict this To understand the issue better, we introduce a richer decision theory that incorporates knowledge into the utility function

In a typical decision problem, we are interested in the results at a particular time T in the future after any decisions have been made and after enough of the state of nature has been learned to determine the impact of our decisions In order to be flexible regarding, among other relevant matters, a "small world" versus a "grand world" framing of a decision problem, we shall let the state of nature ω describe no more than we need to make sense out of each individual decision problem We shall divide the collection of future outcomes into three parts

- The "state of nature" ω will consist of those quantities about which a decision maker (DM) is uncertain at some time during the decision problem The collection of all states of nature is Ω States of nature can include unchanging facts as well as stochastic processes They can include amounts of wealth that might result from gambles They can also include inputs to decisions that might be made at later stages in a decision problem We assume that the state of nature stays the same over time, but our knowledge and/or beliefs about it can change In particular, if the state of nature includes a random variable whose value becomes known later, we interpret this as a change in our knowledge about the random variable rather than a change in the random variable

---

[1] See I J Good, "On the Principle of Total Evidence," *British Journal of the Philosophy of Science*, XVII (1967) 319–21, Howard Raiffa and Robert Schlaifer, *Applied Statistical Decision Theory* (Graduate School of Business Administration, Harvard University, Division of Research, 1961) See also Frank Ramsey, "Weight or the Value of Knowledge," *British Journal for the Philosophy of Science*, XLI (1990) 1–4

- The "knowledge base" is the collection of all information known by the DM If more than one decision maker is being contemplated, each DM has his/her own knowledge base The knowledge base might include the knowledge that certain choices are available now or for future decision problems as well as the various random variables that might be available for observation and when they might become available The knowledge base can, and often will change over time We will denote the knowledge base sequence $\Psi = \{\psi_t \;\; 0 \leq t \leq T\}$, where the subscript t denotes time We use $t = 0$ to stand for "now" when the DM starts thinking about the decision problem, and $t = T$ stands for the time at which all relevant information becomes available and the DM has experienced the impact of the decision The impact of the decision can be cumulative over a time interval, and it is possible that one might want to let $T = \infty$ to indicate that the impact never stops accumulating We assume that, at each time t, the DM has no uncertainty about the corresponding value of the knowledge base ψ_t We do not, however, require that the knowledge base be nondecreasing over time That is, the theory allows that a DM might forget something that was known earlier We will use Ψ with various sub- and superscripts to denote sets of possible values for ψ_t at various times
- The set of acts, gambles, choices, or decisions could easily depend on what information ψ_t the DM has at each time t For example, one cannot choose a decision that explicitly depends on a random variable that one will not have observed at the time that the decision must be implemented So, we denote the set of decisions available at time t when the knowledge base is ψ_t as $\mathcal{D}(\psi_t)$ In a normal-form sequential decision problem, $\mathcal{D}(\psi_0)$ consists of sequences of choices for each stage in the problem where each choice can be a function of what will be known at that stage At time t, $\mathcal{D}(\psi_s)$ for $s > t$ might be random because it could depend on things that might be learned between times t and s In an extensive form decision problem, we treat each stage as a separate decision problem with the same T and the same utility function but with a changing information base sequence

EXAMPLE 1 Suppose that all I care to think about now is where I will be one hour from now Also, I care only to distinguish three possible descriptions of where I will be, namely at work, at home, or elsewhere We could let Ω consist of the three places I could be, and we could let $\mathcal{D}(\psi_0)$ consist of three choices, namely to stay at work, go home, or go somewhere else In this case, we assume that my decision will determine the state of nature with probability 1 At the time that the choice of decision must be made, I cannot have a probability distribution over states because the state is a function of my decision However, suppose that I do not have to decide which action to take for 1/2 hour Then, I could have a marginal probability over Ω for 1/2 hour until it becomes time to decide During that 1/2 hour, I might learn things that help to influence which choice I make

At each time t, let Ψ_t be the set of all possible knowledge bases Then the utility at time t is a bounded function U defined on

$$\Omega \times \bigcup_{\psi_t \in \Psi_t} [\{\psi_t\} \times \mathcal{D}(\psi_t)]$$

That is, the utility at time t is a bounded function of the state, the knowledge base, and the decision, denoted $U(\omega,\psi_t,d)$ As the knowledge base is allowed to (but not required to) accumulate the DM's experiences over time, the utility at time t can depend on what has happened in the past Since the knowledge base is allowed to include anticipated experiences in the future, the utility at time t can depend on what the DM anticipates might happen in the future If $\mathcal{D}(\psi_t)$ is empty, that is, there are no decisions to be made, we shall fill the third argument with the symbol □

EXAMPLE 2 (SIMPLE DECISION PROBLEM) Consider a statistical decision problem with action space $\aleph$, parameter space Θ, and loss function L, where $L(\theta,a)$ stands for the loss that results from choosing action a when θ is the parameter Assume that the loss function does not change with time Assume that no relevant data will be observed by time T, so that the knowledge base is $\psi_t = \psi_0$ for $0 \leq t \leq T$ Let $\mathcal{D}(\psi_0)$ be the set of randomized rules, that is probability distributions over $\aleph$ Assume that $\aleph$ is a nice enough space so that each randomized rule δ can be constructed as the distribution of a function f_δ of a single random quantity R independent of everything else and taking values in a set $\mathcal{R}$ (All Polish spaces are of this type) In particular, a nonrandomized rule $\delta_a(A) = I_A(a)$ for $a \in \aleph$ and $A \subseteq \aleph$ corresponds to $f_{\delta_a}(r) = a$ for all r Let $\Omega = \Theta \times \mathcal{R}$, where the last coordinate will hold the value of R Let Q be the DM's prior distribution over Θ Then, for each nonrandomized choice a, we can denote the utility at time t by $U((\theta,r),\psi,\delta_a) = -cL(\theta,a) + b$ for some $c > 0$ and real b For randomized δ, $U((\theta,r),\psi,\delta) = -cL(\theta, f_\delta(r)) + b$ The expectation of this, over the distribution v of R, involves the usual loss of a randomized rule

$$\int U((\theta,r),\psi,\delta)v(dr) = \int_{\aleph}[-cL(\theta,a)+b]\delta(da) = -cL(\theta,\delta)+b$$

To arrive at this point in the analysis of the problem, one does not make any use of the prior distribution over Θ. Hence, this much of the analysis is equally suitable for both Bayesians and frequentists.

EXAMPLE 3 (STATISTICAL DECISION PROBLEM). Expand Example 2 to include data X taking values in a set $\mathcal{X}$. In this example, let $\mathcal{D}(\psi_0)$ be a set of deterministic functions from $\mathcal{X}$ to $\aleph$ so that only nonrandomized rules are available. Let $\Omega=\Theta\times\mathcal{X}$. In the typical statistical decision problem, we would have

$$U((\theta,x),\psi_t,\delta) = -cL(\theta,\delta(x))+b, \tag{6}$$

for all x and all t, so that utility depends on the data $X = x$ only through the value of $\delta(x)$ and does not change with time. Up to this point, the analysis does not depend on any probability distributions over Θ. Assume now that the DM has a subjective distribution Q over Θ and a regular conditional distribution $P(\cdot|\theta)$ over $\mathcal{X}$ given θ for each $\theta \in \Theta$. Let $Q(\cdot|x)$ represent the posterior distribution over Θ given $X = x$. From (6) it follows that a Bayesian will, for each $x \in \mathcal{X}$, choose $\delta(x)$ equal to that a that minimizes $\int_{\mathcal{X}} L(\theta,a)Q(d\theta|x)$, the usual posterior risk. A non-Bayesian will compute

$$\int_{\mathcal{X}} U((\theta,x),\psi_t,\delta)P(dx|\theta) = -cR(\theta,\delta)+b, \tag{7}$$

where R is the classical risk function. For a maximin solution, one would compute for each δ the infimum of (7) over θ and then choose that δ with the largest infimum.

EXAMPLE 4 (SEQUENTIAL DECISION PROBLEM IN NORMAL FORM). Consider a normal form sequential decision problem in which the decision maker gets to decide how much data to observe. Suppose that the nth observation costs $c_n > 0$ and that the terminal loss is $L(\theta,a)$, where a is the terminal action and $\theta \in \Theta$ is an unknown parameter. To allow for a terminal decision before any data are observed, let $c_0 = 0$. Let X be the potential data sequence taking values in a sequence space $\mathcal{X}$. Let $\mathcal{D}(\psi_0)$ consist of decision rules $\delta = (N,d)$. Here N is either a positive-integer-valued function of data sequences $x \in \mathcal{X}$ such that $\{N = n\}$ is measurable with respect to the first n coordinates of x or $N = 0$ with probability 1 (meaning that no data will be observed). Also, $d = (d_0, d_1, d_2, \ldots)$ is a sequence of functions such that each d_n for $n \geq 1$ is a function of the first n coordinates of x and takes values in $\aleph$ and d_0 is some element of $\aleph$. The interpretation of such a rule is the following. If the DM decides to observe no data, d_0 will be the terminal action. Otherwise the DM observes $N(X) > 0$ observations and takes terminal action $d_{N(X)}(X)$. Now, $\Omega=\Theta\times\mathcal{X}$. To express the usual sequential decision problem in the present framework, we assume that

$$U((\theta,x),\psi_i,\delta) = -c\left[L(\theta, d_{N(x)}(x)) - \sum_{n=0}^{N(x)} c_n\right] + b,$$

for some real b and $c > 0$ Once again, we have assumed that utility depends on the value of $X = x$ only through $\delta(x)$

EXAMPLE 5 (SEQUENTIAL DECISION PROBLEM IN EXTENSIVE FORM) Consider the same general situation as in Example 4, but this time, assume that the DM gets to reconsider her choice after each coordinate of X is observed That is, the sample size N need not be a predetermined measurable function of X such that $\{N = n\}$ is a function of the first n coordinates of X. It is more convenient to express an extended form problem as a sequence of one-stage decision problems in which the utility at each stage still depends on the utilities at other stages To be precise, let t_0 be now and let $t_1 < t_2 <$ be a sequence of times at which additional data might be collected depending on what decisions are made at earlier times Of course $t_0 < t_1$ At each time t_i ($i = 0,1,$) at which a terminal decision has not yet been made, the DM has the option of choosing to observe more data X_{i+1} at time $i + 1$ or of making a terminal decision at the present time If we let N denote the number of observations that will eventually be chosen, then N is random until a terminal decision is made Similarly, we will let A denote the terminal action eventually chosen so that A is random until that time when a terminal action is finally chosen For the decision problem at each time t_i, the set of possible decisions available is either empty (if a terminal decision was made in an earlier decision problem) or it consists of the terminal actions $\aleph$ plus the decision to continue sampling Using the notation of Example 4, the utility for the decision problem at each time t_i prior to a terminal decision is

$$U((\theta,x),\psi_i,\delta) = \begin{cases} -c\left[L(\theta,a) - \sum_{n=0}^{i} c_n\right] + b & \text{if } \delta \text{ chooses terminal action } a, \\ -c\left[L(\theta,A) - \sum_{n=0}^{N} c_n\right] + b & \text{if } \delta \text{ chooses to observe more data} \end{cases}$$

In the framework of Examples 4 and 5, the pre-data value of a particular decision rule is the same whether the decision problem is put into normal form or into extensive form In the normal form there is only one decision point among rules and each rule involves potentially many contingent directions for what to do based on the outcomes of subsequent observations In extensive form, there are many decision points, each one corresponding to a change in available information Bayesian analysis yields the same assessment for a decision rule regardless in which form the decision is cast However, for the decision theories discussed in section v, where uncertainty is represented by sets of probability distributions, the two forms are not equivalent[2]

III THE COST AND VALUE OF INFORMATION

Section I discusses the value of cost-free information without formally defining cost-free information To assess the cost of observing some information or the value of the information, one needs to be able to imagine the possibility of observing (and possibly using) the information as well as the possibility of not observing it Hence, we set up decision problems in which there is at least one stage at which the DM can choose whether or not to observe the information Since the cost and/or value of information can change over time (imagine observing a relevant sample after having to make a decision), we attach a time of observation to any definition of cost or value Furthermore, the cost or value of information is relative to a specific decision problem

The value of any information depends on what one does with it, so any definition will have to take into account how information is used For example, in a Bayesian analysis, one tries to maximize expected utility, while others might try to maximize the infimum of the utility function We will assume that the DM wants to maximize some nondecreasing functional h of the utility function Examples of h include the integral with respect to a probability over ω and the minimum over all ω Different functionals h correspond to different theories of decision making To be precise, let h be a real-valued function defined on the space of all bounded functions $g\ \ \Omega \to \Re$ that satisfies $h(g_1) \leq h(g_2)$ whenever $g_1(\omega) \leq g_2(\omega)$ for all ω We assume that the DM with knowledge base sequence ψ chooses a decision rule δ in order to maximize $h(U\ (\ ,\psi_t,\delta))$ for some time t There are principles of decision making that do not fit this description, but many popular ones do

To distinguish cost from value, we think of the net value of information as some measure of the change in utility one would achieve by observing the information and using it however one saw fit as opposed to not observing it We would like to define the cost to be the change in utility one would achieve by observing the information but then ignoring it in making decisions. This is problematic because it may not be possible to ignore some information after observing it. For this reason, we take a less ambitious approach and merely define cost-free information in section III but do not quantify the cost of information that is not cost-free.

---

[2] See Teddy Seidenfeld, "When Normal and Extensive Form Decisions Differ," in Dag Prawitz, Brian Skyrms, and Dag Westerstahl, eds , *Logic, Methodology and Philosophy of Science IX* (New York Elsevier, 1994), pp 451–63, for in-depth discussion of how normal and extensive form decision problems differ

III.1. The Net Value of Information. Suppose that a DM might be able to observe some information X at a time $t_1 \geq 0$ and is faced with a decision problem as described earlier. The DM wants to say how valuable is the information X at time t_1 in the decision problem. To do this, we embed two versions of the original decision problem in a 2-stage problem in normal form (called the extended decision problem) in which the first stage involves a decision at time 0 of whether or not to observe X at time t_1. To be clear, we imagine the DM deciding at time 0 whether or not to observe X, but the actual observation of X may occur later at time t_1. The second stage of the problem will be one of the two versions of the original decision problem, one in which X is observed (version 1) and one in which X is not observed (version 0). The versions of the original decision problem may themselves be sequential, but this will be immaterial for the current discussion. The set of decision rules available in the extended decision problem at time $t = 0$ is $\mathcal{D}(\psi_0) = (N, d_0, d_1)$ where $N = 0$ means to not observe X, $N = 1$ means to observe X, and d_i is a decision rule for version $i (i = 0,1)$. This matches the notation used in Example 4. To make it easier to discuss the two versions together, let $\psi = \{\psi_t\}_{t=0}^{T}$ be a knowledge base for which it is known that X is not to be observed at time t_1, and let $\psi^* = \{\psi_t^*\}_{t=0}^{T}$ denote the alternative knowledge base in which it is known that X will be or has been observed at time t_1. We assume that ψ_0^* is the same as ψ_0 aside from the knowledge that X will be observed, so that the two versions start with essentially the same information before the decision is made whether or not to observe X.

We assume that the DM is able to determine what she would do if X were observed and what she would do if X were not observed. Formally, we assume that both version 0 and version 1 of the original decision problem can be solved by the DM. This could be possible even in cases in which the DM already knows that she will observe X but still knows what she would have done if X were not to be observed. The net value of the information X is then defined to be the difference

$$h(U(\,,\psi_t^*,\delta_1)) - h(U(\,,\psi_t,\delta_0)),$$

where δ_i is the decision rule that the DM would use if forced to solve version i of the original decision problem for $i = 0, 1$ at time t.

EXAMPLE 6 (MAXIMIN DECISION MAKER) Suppose that the DM chooses decision rules using a maximin criterion To be more specific, consider a traditional decision problem such as the ones described in Example 3 where U is given by (6) Suppose that a decision must be made at time $t \geq t_1$ A maximin decision maker, in such a problem, uses the functional

$$h(g) = \inf_{\theta \in \Theta} \int_X g(\theta,x) P(dx|\theta) \tag{8}$$

Substituting the utility from (6) into (8) we obtain

$$h(U(\,,\psi_t^*,\delta)) = \inf_{\theta \in \Theta} \int_X [-cL(\theta,\delta(x)) + b] P(dx|\theta),$$

for each rule $\delta \in \mathcal{D}(\psi_t^*)$ In version 0, all decision rules δ have to be constant as a function of x Hence, we have

$$h(U(\,,\psi_t,\delta)) = \inf_{\theta \in \Theta} [-cL(\theta,\delta) + b],$$

for each $\delta \in \mathcal{D}(\psi_t)$ So, the net value of X would be

$$\sup_{\delta \in D(\psi_t^*)} \inf_{\theta} (-cR(\theta,\delta) + b) - \sup_{\delta \in D(\psi_t)} \inf_{\theta} (-cL(\theta,\delta) + b)$$

where $R(\theta,\delta)$ is the classical risk function

$$R(\theta,\delta) = \int_X L(\theta,\delta(x)) P(dx|\theta)$$

As a specific example, consider the following augmentation of the example presented by Leonard Savage[3] in which we give a specific data distribution Let $\Omega = \Theta \times \mathcal{X}$ where the parameter space $\Theta = \{1, 2\}$ and the sample space $\mathcal{X} = 1, 2, 3$ Let the space of terminal decisions $\aleph = \{1,2\}$ At time t_0 the DM must decide whether to observe X taking values in $\mathcal{X}$ At time $t_1 > t_0$, the DM has to choose one of the two terminal actions from $\aleph$ Let X have the following conditional distribution given θ

$$\Pr(X = i | \theta = j) = \begin{cases} 0\,5 & \text{if } i = j = 1 \text{ or } i = j = 2 \text{ or } i = 3, \\ 0 & \text{otherwise} \end{cases}$$

The random variable X identifies θ with probability 0 5 (when $X \neq 3$) and it is irrelevant with probability 0 5 (when $X = 3$) Let the utility have the form of (6) with $c = 1$, $b = 0$ and consider the following loss function regardless of whether X is observed

---

[3] Savage, *The Foundations of Statistics* (New York Wiley, 1954), p 170

$$L(\theta,a)=\begin{cases} 1 & \text{if } a = 1, \\ 10 & \text{if } \theta = 1, a = 2, \\ -1 & \text{if } \theta = 2, a = 2 \end{cases}$$

That is,

$$U((\theta,x),\psi_t,\delta)=-L(\theta,\delta(x)),$$

for all t, θ, and x Intuition suggests that X has value for the DM unless θ is known with certainty If X is not observed, then the maximin decision is to choose terminal action $a = 1$ In this case $h(U(\,,\psi_t,1)) = -1$ If X is to be observed, the maximin rule is any Bayes rule with respect to the least favorable prior The least favorable prior is $\Pr(\theta=1)=1$, and the collection of Bayes rules is the set of rules δ_α where, for each α, $\delta_\alpha(1) = \delta_\alpha(3) = 1$ and $\delta_\alpha(2)$ randomizes between 1 and 2 with $\Pr(\delta_\alpha(2) = 1) = \alpha$ for arbitrary $\alpha \in [0,1]$ The risk functions of these rules are

$$R(\theta,\delta_\alpha)=\begin{cases} 1 & \text{if } \theta = 1, \\ \alpha & \text{if } \theta = 2 \end{cases}$$

For all of these rules, $h(U(\,,\psi_t,\delta_\alpha)) = -1$, so the maximin decision maker assigns 0 net value to the information X, contrary to intuition

Savage claims that maximin decision making is "utterly untenable for statistics" based on the fact that the seemingly relevant data in Example 6 has 0 value (*ibid*, p 170) (We will see in section iii 2 that the X in that example is cost-free) Savage contrasts the maximin rule with maximin-regret, which we describe next

EXAMPLE 7 (MAXIMIN-REGRET DECISION MAKER) A variation on maximin decision making is to shift the utility function so that the maximum value is the same in every state, typically 0 In symbols, replace U by

$$U'(\omega,\psi_t,\delta)=U(\omega,\psi_t,\delta)-\max_{\delta} U(\omega,\psi_t,\delta) \tag{9}$$

The maximin-regret decision maker then behaves the same as a maximin decision maker whose utility is U' instead of U In particular, for the calculations that appear in Examples 3 and 6, we replace $L(\theta,\,)$ by $L(\theta,\,) - \min_a L(\theta,a)$ everywhere and set $b = 0$ wherever it occurs The functional h is still (8)

Now, reconsider the example of Savage (*ibid*, p 170) from this point of view The modified loss function is

$$L'(\theta,a)=\begin{cases} 9 & \text{if } \theta = 1,\ a = 2, \\ 2 & \text{if } \theta = 2,\ a = 1, \\ 0 & \text{otherwise} \end{cases}$$

If the DM does not observe X, the maximin-regret decision is the randomized rule δ that chooses $a = 1$ with probability 9/11 and chooses $a = 2$ with probability 2/11 For this rule, $h(U(,\psi_t 0,\delta) = -18/11$ However, if X is observed, the risk function of the general randomized rule δ_β that chooses $a = 1$ with probability β_i when $X = i$ is observed equals

$$R'(\theta,\delta_\beta) = \begin{cases} \frac{9}{2}(2-\beta_1-\beta_2) & \text{if } \theta = 1, \\ \beta_2+\beta_3 & \text{if } \theta = 2 \end{cases}$$

The maximin rule corresponds to $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = 9/11$ which has $h(U(,\psi_t,\delta_\beta)) = -9/11$ So, the maximin-regret decision maker assigns net value $(-9/11) - (-18/11) = 9/11$ to the information X

Savage suggests that maximin-regret decision making is not as prone to the objection of assigning zero value to relevant data (*ibid*, p 200) However, Giovanni Parmigiani[4] shows that there are examples in which maximin decision making gives positive net value to data that are assigned zero net value by maximin-regret

EXAMPLE 8 (BAYESIAN DECISION MAKER) Suppose that the DM wants to choose the decision rule δ at time t that maximizes expected utility Let the DM's distribution (now) for ω be P Then $h(g) = \int_\Omega g(\omega)P(d\omega)$ and

$$h(U(,\psi_t,\delta)) = \int_\Omega U(\omega,\psi_t,\delta)P(d\omega)$$

Next, consider the special case of Example 3 in which $\omega = (\theta, x)$ Then the net value of the information X is the difference between the extreme terms in (5) if the following two conditions hold

- the utility has the simple form $U((\theta, x),\psi_t,\delta) = U(\delta(x),\theta)$ for all t, and
- (X, θ) has a joint density $p(x, \theta)$ with respect to Lebesgue measure

In Savage's example, suppose that the Bayesian DM starts with $\Pr(\theta = 1) = p$ Without observing X, it is easy to see that a Bayesian DM will choose $a = 1$ if $p > 2/11$ and will choose $a = 2$ if $p < 2/11$ Either action can be chosen if $p = 2/11$ Call this Bayes decision δ_p Then

$$h(U(,\psi_t,\delta_p)) = \begin{cases} 1-11p & \text{if } p < 2/11, \\ 1 & \text{if } p \geq 2/11 \end{cases} \tag{10}$$

If X is observed, then the Bayesian DM chooses $a = 1$ if $X = 1$ and chooses $a = 2$ if $X = 2$ If $X = 3$, the DM uses the same δ_p she would have used if X had not been observed For this rule δ_p^*, $h(U(\,,\psi_t,\delta_p^*))$ equals one-half of (10) (corresponding to $X = 3$) plus $0\ 5 - p$ (corresponding to $X \neq 3$) The net value of X for the DM who believes $\Pr(\theta = 1) = p$ a priori is

---

[4] Parmigiani, "Minimax, Information, and Ultrapessimism," *Theory and Decision*, XXXIII (1992) 241–52

$$\begin{cases} \frac{9}{2}p & \text{if } p < 2/11 \\ 1-p & \text{if } p \geq 2/11 \end{cases} \tag{11}$$

Notice that the largest possible net value occurs at $p = 2/11$ and equals 9/11, the same net value assigned by the maximin DM The least favorable distribution in the maximin-regret problem is $\Pr(\theta = 1) = 2/11$ for both version 0 and version 1

EXAMPLE 9 (Γ-MAXIMIN DECISION MAKER) We will discuss Γ-maximin decision rules in more detail in section v For now, we can understand Γ-maximin decision making as a variant on Bayesian decision making in the following sense One replaces the single distribution P over states used by the Bayesian decision maker in Example 8 by a closed convex set $\mathcal{P}$ of distributions Then, one ranks decision rules by

$$h(U(\,,\psi_t,\delta)) = \inf_{P\in\mathcal{P}} \int_\Omega U(\omega,\psi_t,\delta)P(d\omega) \tag{12}$$

In Savage's example, suppose that the Γ-maximin DM chooses $\mathcal{P}=[p_1,p_2]$ If $p_1 \leq 2/11 \leq p_2$, then the DM behaves just like the maximin-regret DM because the infs in (12) will occur at the least-favorable distribution in both version 0 and version 1 If $p_2 < 2/11$, the DM will behave like the Bayesian in Example 8 with $p = p_2$ If $p_1 > 2/11$, the DM will behave like the Bayesian with $p = p_1$ The net value of X can be found for these last two cases in (11)

III 2 Cost-Free Information This section formally defines what we mean by saying that observing information X at a specific time t_1 is cost-free We will do this in a manner that does not depend on the decision principles to which one adheres In order for information at a specific time to be cost free for every mode of decision making, we require several conditions First, we require that it is possible not to observe X at time t_1 Second, we require that the information be *decision-theoretically ignorable*, meaning that every decision rule that is available without the information is also available with the information so that one could ignore the information when choosing a decision rule To make this more precise, let $\psi = \{\psi_t\}_{t=0}^T$ be a knowledge base sequence for which it is known that X is not to be observed at time t_1, and let $\psi^* = \{\psi_t^*\}_{t=0}^T$ denote the alternative knowledge base sequence in which it is known that X will be or has been observed at time t_1 To say that observing X at time t_1 is decision-theoretically ignorable, we mean that for each decision time t and each $\delta \in \mathcal{D}(\psi_t)$ there is a $\delta^* \in \mathcal{D}(\psi^*_t)$ which does exactly what δ would do under all states of nature that are still possible, essentially ignoring X to the extent possible To put it another way, learning X at time t_1 does not eliminate any decision rules from consideration There is a subtle issue that should be addressed at this point Suppose that, even if the DM does not observe X at time t_1, she will nevertheless observe it at some later time A simple example will illustrate the issue

EXAMPLE 10 Let $\aleph$ consist of two points $\aleph = \{a_0, a_1\}$ Let X be a Bernoulli random variable that we believe to be independent of everything that is relevant in this decision problem Let $t_1 = 0$, and let $0 < t_2$ Suppose that X will be observed by time t_2 under all circumstances, but we have the option of observing it at time t_1 Suppose also that, at time t_2, we must choose between two decision rules $\delta_0 \equiv a_0$ and $\delta_1(X) = a_X$ That is, we can either pick action a_0 by choosing δ_0 or we can choose δ_1 which randomizes between a_0 and a_1 by using X if $X = 0$ then we will pick a_0 and if $X = 1$ we will pick a_1 So, the terminal action is not finally decided until time t_2 when X will definitely be known If we choose to observe X at time t_1 and $X = 0$ is observed, then the only terminal action that will be available to us is a_0 regardless of whether we choose decision rule δ_0 or δ_1 Nevertheless, the decision rule δ_1 is still available at time t_2 It just happens that we already know that $\delta_1(X) = a_0$

In Example 10, X is decision-theoretically ignorable because the decision rule δ_1 is still available when it comes time to choose regardless of the fact that we might already know that it will make the same choice as δ_0 When it comes time to make our decision, we are in the same situation regardless of whether we observed $X = 0$ at time t_1 or at some later time

Information X that is decision-theoretically ignorable will be called *cost-free* if, for every state of nature ω, every knowledge base sequence ψ, every time t, and every $d \in \mathcal{D}(\psi_t)$,

$$U(\omega, \psi_t, \delta) = U(\omega, \psi^*_t, \delta^*), \tag{13}$$

where δ^* was defined above as the decision rule in $\mathcal{D}(\psi^*_t)$ that makes the same choices as δ under all circumstance In other words, if the DM ignores X, she will achieve the same utility as if X had not been observed In particular, there can be no fee for merely observing X.

EXAMPLE 11 (RETURN TO EXAMPLE OF SAVAGE) This example is set up in the second half of Example 6 In the notation of this section, let $\mathcal{D}(\psi_{t_1})$ contain all randomized decision rules that are just probabilities over $\aleph = \{1,2\}$, and let $\mathcal{D}(\psi^*_{t_1})$ consist of all randomized decision rules that are functions of X It is easy to see that for each rule δ_p in

$\mathcal{D}(\psi_{t_1})$ that chooses $a = 1$ with probability p and chooses $a = 2$ with probability $1 - p$, there is a $\delta_p{}^* \in \mathcal{D}(\psi^*_{t_1})$ such that $\delta^*_p(x) = \delta_p$ for each $x = 1, 2, 3$ Hence, X is decision-theoretically ignorable It is also easy to see that $U(\omega,\psi_t,\delta_p) = U(\omega,\psi^*_t,\delta^*_p)$ for all p, ψ, and t Hence, the information X is cost-free in this example

Now return to the example at the start of section II

EXAMPLE 12 (THE TAXI DRIVER) Recall that I hold a ticket to a mystery play and take a taxi to the theater The taxi driver knows who-done-it and offers to inform me unless the tip is sufficiently large Does it make sense to pay to avoid learning this information? Assume that I am committed to attending the mystery so that I need not consider the possibility of doing something else This makes the information offered by the taxi driver decision-theoretically ignorable because I have no decision points after I choose the tip For this reason, we will write the utility as $U(\omega,\psi_t,\square)$ to indicate that no decisions are available in either version 0 or version 1 of the problem If we determine that the information is not cost-free, then we are faced with the extended decision problem of whether to pay not to observe it For simplicity, suppose that I am contemplating, at time 0, only two different tips, 0 and $v > 0$, where v is sufficiently large to prevent the taxi driver from telling me who-done-it At present, I believe that attending the mystery will have a random (that is, unknown) effect on several of my emotions at certain times in the future with various probabilities that reflect my opinion of how well the mystery is produced and what I know at the time of the play In particular, the play offers the possibility of suspense and surprise If I do not know who-done-it, the probabilities of high levels of suspense and surprise are larger than if I do know Similarly, if I do not know who-done-it, the probability of boredom is smaller than if I do know A poorly written/produced play might cause boredom even if I do not know who-done-it We can interpret the information about who-done-it to be data that might be added to my knowledge base at or before time T, when the mystery will be finished The taxi driver has offered me the choice of which tip to give, which is assumed to be equivalent to whether or not to obtain the information of who-done-it No further decision point is available after this decision is made

To be explicit, let $\Omega = \{1, \quad ,\ell\} \times \Upsilon$, where ℓ is the number of characters in the play, and Υ is a space each of whose elements specifies a set of values for the various aspects of the quality of the play Suppose that I have a distribution μ over the space Υ that gives my opinion of the play Let ψ_0 be my current knowledge base, and assume that I will not learn anything else before I have to choose a tip at time t_1, so that $\psi_{t_1} = \psi_0$ For each $i = 1, \quad ,\ell, t > t_1$, and $x > 0$ let $\psi_{t,i,x}$

be the augmented knowledge base at time t assuming that I learned that character i done-it before attending and assuming that I have paid x as a tip Let $\psi_{t,x}$ be the corresponding knowledge base if I do not learn who-done-it but pay a tip of x To avoid certain philosophical difficulties, suppose that, prior to getting in the taxi I have already considered the possibility that I might learn who-done-it before seeing the play and that I had already considered the possibilities of giving various tips even if I had never dreamed that the two might be related Assume, for simplicity, that neither $U((i,y), \psi_{t,x},\square)$ nor $U((i,y),\psi_{t,i,x},\square)$ depends on i for fixed $y \in \Upsilon$ and fixed x These assumptions merely mean that I have no particular interest in which character did it Fix $t > t_1$ and $x > 0$ and let $U_x=\int_{\Upsilon} U((i,y),\psi_{t,i,x},\square)\mu(dy)$ and $U^x=\int_{\Upsilon} U((i,y),\psi_{t,x},\square)\mu(dy)$ These are respectively the expected utilities that I would have calculated at time t after giving a tip of x under the assumptions of learning and not learning who-done-it before seeing the play

It makes sense that $U_x < U^x$ for all x, that is, it is better not to know who-done-it, regardless of what tip I might give Also, $U_x > U_w$, and $U^x > U^w$ whenever $x < w$, meaning that my utility goes down with the size of the tip, both in the case of learning and of not learning who-done-it The expected utility of choosing tip v is then Uv while the expected utility of choosing tip 0 is U_0 The fact that $U_v < U^v$ makes the information not cost-free Which tip we should give hinges on whether or not $U^v > U_0$, the only comparison not fixed so far by the problem description This comparison is between the disutility of the large tip v, on the one hand, to the disutility of ruining the experience of the play on the other

Examples 4, 5, and 12 are cases in which the data are not cost-free

IV THE VALUE OF COST-FREE INFORMATION

Assume that X is cost-free Then a DM who maximizes expected utility will not want to pay to avoid learning X To see this, argue as follows Let $P(dx,d\omega)$ stand for the joint distribution of X and ω Even though we have assumed that X is a function of ω, this joint distribution still makes sense as a measure Also, let $P(d\omega \mid x)$ stand for the posterior distribution of ω given $X = x$ Let δ^* be the decision that the DM would make without observing X, and let $\delta^{*\prime}$ be the corresponding rule that observes X but ignores it Then for each t

$$\int_X \int_\Omega U(\omega,\psi_t,\delta^*)P(dx,d\omega) \tag{14}$$

is the DM's expected utility at time t without observing X If $X = x$ were observed, the DM would choose the decision δ to maximize

$$\int_\Omega U(\omega,\psi_{t,x},\delta)P(d\omega \mid x) \tag{15}$$

For each x, we use (13) to generalize (4) as

$$\max_{\delta} \int_{\Omega} U(\omega,\psi_{t,x},\delta)P(d\omega|x) \geq \int_{\Omega} U(\omega,\psi_{t,x},\delta^{*\prime})P(d\omega|x) = \int_{\Omega} U(\omega,\psi_{t},\delta^{*})P(d\omega|x) \tag{16}$$

The rest of the argument is virtually identical to that in section I.

The result above (as well as the one in section I) applies only to joint probability distributions on the data and state of nature that are countably additive and proper, as the following examples show. Both of these examples assume that the available data are cost-free, that is, the utility function is the same before and after observing the data. What fails from the result above and that of section I is that finitely additive and improper joint distributions do not always factor into marginals and conditionals the way that countably additive joint distributions do.

EXAMPLE 13 (FINITELY, BUT NOT COUNTABLY ADDITIVE DISTRIBUTIONS). Suppose that there are two states of the world, A and B, each of which has probability 0.5 in your current opinion. Imagine that you can observe a positive integer N. If A is true, the integer N has a geometric distribution, as follows:

$$P(n|A) = (1/2)^n,$$

for each integer $n > 0$. However, if B is the case, the integer N is uniformly distributed on the integers in your opinion. There are many such finitely additive distributions,[5] but they all have the property that

$$P(n|B) = 0$$

for each integer $n > 0$.

If a particular integer, say $N = 3$, is observed, an easy application of Bayes's Theorem shows that

$$P(A|N=3) = 1$$

and in fact, this is true whatever value n of N is observed. Hence you are in the peculiar state of belief that although your prior is even between A and B, you know that conditional on the observation of N, regardless the value of N, you believe with certainty that A is true.

---

[5] See Joseph B. Kadane and Anthony O'Hagan, "Using Finitely Additive Probability Uniform Distributions on the Natural Numbers," *Journal of the American Statistical Association*, XC (1995): 626–31; also Oliver Schirokauer and Kadane, "Finitely Additive Distributions on the Natural Numbers: Shift Invariance," *Journal of Theoretical Probability*, XX, 3 (2007): 429–41.

KP 2725

Which then is your prior, what you believe now, or what you know you would believe if you could observe N?[6]

Suppose that you currently hold a ticket that pays $1 if B is true, and −$1 if A is true. Currently your expected winnings are $0. However, you know that if you were to obtain the integer N, your expected winnings would be −$1. Would you pay $0.50 not to receive the data? It seems that you would have to.

Where does the proof given above fail for finitely but not countably additive distributions? The proof fails because conditional probability does not work the same way for finitely additive and countably additive distributions. Technically, it concerns the failure of conglomerability,[7] that is

$$P(A|N=n)>P(A) \text{ for all } n > 0.$$

EXAMPLE 14 (IMPROPER DISTRIBUTIONS). There are Bayesians who use improper distributions (that is, those that integrate to infinity) as a way of modeling "ignorance"; sometimes these priors are called "reference" or "objective." The simplest case of these arise as apparent results of limits of proper priors. Suppose for example that the datum x is normally distributed with mean W and precision (inverse variance) 1. Suppose also that W has a normal prior with mean 0 and precision r. Then it is well known that the posterior distribution of W given x is again normal with mean the precision-weighted average of x and 0, and precision equal to $r + 1$. Now imagine that r approaches 0, so the variance of the prior goes to infinity. Then the posterior of W given x approaches a normal distribution with mean x and precision 1. This is the same answer that would have been obtained from using an improper uniform distribution on W, and doing a formal calculation of the posterior. The hope is that because the improper distribution is the limit of proper posteriors, nothing "bad" happens in the limit.

Now suppose that the precision R of the normal observation is not known, but that the conjugate prior is imposed, that is, W given $R = r$ is normal with mean 0 and precision τr, and R has a $\Gamma(\alpha, \beta)$ distribution with $\alpha > 1$. The general form of the posterior, given $X = x$ is then[8] as follows: the posterior on W given $R = r$ is normal as before, with mean the precision-weighted average of x and 0 with precisions r and τ respectively, and the posterior of R is $\Gamma(\alpha + 1/2, \beta + \tau x^2/[2(\tau + 1)])$ As τ approaches zero, the conditional posterior distribution of W given $R = r$ is as before, normal with mean x and precision r, and the posterior of R is $\Gamma(\alpha + 1/2, \beta)$

---

[6] We explored this phenomenon in Kadane, Mark Schervish, and Seidenfeld, "Reasoning to a Foregone Conclusion," *Journal of the American Statistical Association*, xci (1996): 1228–35.

[7] For more on this, see Bruno de Finetti, *Theory of Probability*, Volume 1 (New York: Wiley, 1974).

[8] See Morris Herman DeGroot, *Optimal Statistical Decisions* (New York: McGraw-Hill, 1970), pp. 167ff.

Now, we set up a decision problem in which one would pay to avoid learning X. Let $\sigma^2 = 1/R$, the usual variance parameter In the notation used earlier, with $\omega = (W, R)$, let $U(\omega,\psi_t,\delta) = \sigma^2$ Note that the utility depends only on the value of ω, not on any decisions In this case, before any data are observed,

$$E[U(\omega,\psi_0,\delta)] = E(\sigma^2) = \beta/(\alpha - 1), \tag{17}$$

and after $X = x$ is observed say at time t_1,

$$E[U(\omega,\psi_{t1},\delta)] = (\sigma^2 | X = x) = \frac{\beta}{\alpha - 1/2},$$

the same for all x, which is strictly smaller than (17)

A similar result holds no matter how many observations one contemplates For example, suppose that $X = (X_1, \quad ,X_n)$ is a vector of n observations that are conditionally independent and identically distributed given (W, R) with the same distribution described above In this case, (1) is the same as above, namely $\beta/(\alpha - 1)$ With n observations, the posterior distribution of R derived from the improper prior is $\Gamma(\alpha+n/2, \beta+\sum_{i=1}^{n}(x_i-\bar{x})^2/2)$, where $\bar{x} = \sum_{i=1}^{n} x_i/n$ Using the same utility $U(\omega,\psi_t,\delta) = \sigma^2$, we see that (2) equals

$$\frac{\beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2}{\alpha + \frac{n}{2} - 1} \tag{18}$$

To calculate (3), recall that the distribution of X is defined by first conditioning on (W, R) and then integrating with respect to the distribution of (W, R) The integral of (18) with respect to the conditional distribution of X given $(W, R) = (w, r)$ is easily computed as

$$\frac{\beta + \frac{(n - 1)\sigma^2}{2}}{\alpha + \frac{n}{2} - 1}$$

Integrating this with respect to the marginal distribution of (W, R) yields

$$\frac{\beta + \frac{1}{2}\frac{(n - 1)\beta}{\alpha - 1}}{\alpha + \frac{n}{2} - 1},$$

which is then (3) in this example. A little algebra shows that (3) is still smaller than (1). Indeed, the difference is $-\beta/[2(\alpha-1)(\alpha-1+n/2)]$. One would pay, say, half of this amount to avoid seeing the data.

The issue in Example 14 is that the posterior mean of σ^2 differs from the prior mean by an amount that has negative prior mean. That is, even before collecting the data, we expect the posterior mean to be smaller than the prior mean. Just as in Example 13, it is fair to ask what your prior mean of σ^2 really is, what you think now or the mean of what you know you would think if you were to observe X. Just as before, you can be confronted with the prospect of paying not to see the data X, because it would change your mean of σ^2 in a way unfavorable to you. Hence even in the apparently simple case of a sample of observations from a univariate normal distribution, with conjugate priors and flattening on the mean, the same issue arises of potentially paying not to see the data.

V. GENERALIZATIONS OF BAYESIAN DECISION THEORY INVOLVING SETS OF PROBABILITIES

There is recent interest in decision theories involving sets of probabilities. The intuitive idea behind them is to relax the requirement of knowing one's opinion about $p(x, \theta)$, but doing so in a way different from the "reference" - "objective" - "non-informative" prior school. We study three of the leading such theories here. In all of this discussion, the joint distributions will be countably additive and all data will be cost-free.

Of the three decision rules we discuss, perhaps the most familiar one is Γ-Maximin.[9] This rule requires that the decision maker ranks a gamble by its lower expected value, taken with respect to a closed, convex set of probabilities, $\mathcal{P}$, and then to choose an option from $\mathcal{A}$ whose lower expected value is maximum.[10] The Γ-Maximin decision rule creates a preference ranking of options independent of the alternatives available in A: it is context independent in that sense. But

---

[9] When outcomes are cast in terms of a (statistical) loss function, the rule is then Γ-Minimax: rank options by their maximum expected risk and choose an option whose maximum expected risk is minimum.

[10] This decision rule was given a representation in terms of a binary preference relation over Anscombe-Aumann horse lotteries by Itzhak Gilboa and David Schmeidler, "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, XVIII (1989): 141–53, has been discussed by, for example, James O. Berger, *Statistical Decision Theory and Bayesian Analysis* (New York: Springer, 1985, 2nd ed.), section 4.7.6. And recently by Peter D. Grunwald and A. Philip Dawid, "Game Theory, Maximum Entropy, Minimum Discrepancy, and Robust Bayesian Decision Theory" (Technical Report 223, University College, 2002) who defend it as a form of Robust Bayesian decision theory.

Γ-Maximin corresponds to a preference ranking that fails the so-called (von Neumann-Morgenstern's) "Independence" or (Savage's) "Sure-thing" postulate of SEU (Subjective Expected Utility) theory

The second decision rule that we consider, called *E*-admissibility ('*E*' for "expectation"), was formulated by Isaac Levi [11] *E*-admissibility constrains the decision maker's admissible choices to those gambles in A that are Bayes for at least one probability $P \in \mathcal{P}$ That is, given a choice set A, the gamble f is *E*-admissible on the condition that, for at least one $P \in \mathcal{P}$, f maximizes subjective expected utility with respect to the options in $\mathcal{A}$ [12] Savage[13] defends a precursor to this decision rule in connection with cooperative group decision making *E*-admissibility does not support an ordering of options, real-valued or otherwise, so that it is inappropriate to characterize *E*-admissibility by a ranking of gambles independent of the set $\mathcal{A}$ of feasible options However, the distinction between options that are and are not *E*-admissible does support the "Independence" postulate For example, if neither option f nor g is *E*-admissible in a given decision problem $\mathcal{A}$, then the convex combination, the mixed option $h = \alpha f \oplus (1 - \alpha) g (0 \leq \alpha \leq 1)$ likewise is *E*-inadmissible when added to $\mathcal{A}$ This is evident from the basic SEU property the expected utility of a convex combination of two gambles is the corresponding weighted average of their separate expected utilities, hence, for a given $P \in \mathcal{P}$ the expected utility of the mixture of two gambles is bounded above by the maximum of the two expected utilities The assumption that neither of two gambles is *E*-admissible entails that their mixture has *P*-expected utility less than some *E*-admissible option in $\mathcal{A}$

The third decision rule we consider is called *Maximality* by Peter Walley,[14] who appears to endorse it (*ibid* , p 166) If a gamble $f \in \mathcal{A}$ is such that there is no other gamble $g \in \mathcal{A}$ strictly preferred to f, then f is admissible by *Maximality* That is, f is a *Maximal* choice from $\mathcal{A}$ provided that there is no other element $g \in \mathcal{A}$ that, for each $P \in \mathcal{P}$, carries greater expected utility than f does *Maximality* (under differ-

---

[11] Levi, "On Indeterminate Probabilities," this JOURNAL, LXXI, 13 (July 18, 1974) 391–418, and Levi, *The Enterprise of Knowledge* (Cambridge MIT, 1980)

[12] Levi's decision theory is lexicographic, in which the first consideration is *E*-admissibility, followed by other considerations, for example, what he calls a Security Index Here, we attend solely to *E*-admissibility

[13] Savage, *op cit*, section 7 2 Savage's analysis of the decision problem depicted by his Figure 1, p 123, and his rejection of option *b*, p 124 is the key point

[14] Walley, *Statistical Reasoning with Imprecise Probabilities* (New York Chapman and Hall, 1990) There is, for our discussion here, a minor difference with Walley's formulation of maximality involving null-events Walley's notion of maximality requires, also, that an admissible gamble be classically admissible, that is, not weakly dominated with respect to state-payoffs

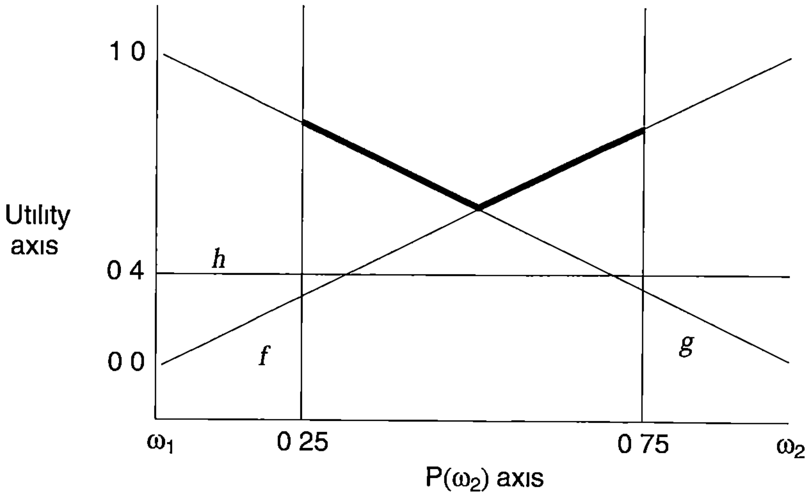


Figure 1. Expected utilities for three acts in Example 15. The thicker line indicates the surface of Bayes solutions.

ent names) has been studied, for example, in Hans G. Herzberger,[15] Levi,[16] Seidenfeld,[17] and Amartya Sen.[18] Evidently, the *E*-admissible gambles in a decision problem are a subset of the Maximally admissible ones.

The three rules have different sets of admissible options. Here is a heuristic illustration of that difference.

EXAMPLE 15. Consider a binary-state decision problem, $\Omega = \{\omega_1, \omega_2\}$, with three feasible options. Option *f* yields an outcome worth 1 utile if state ω_1 obtains and an outcome worth 0 utiles if ω_2 obtains. Option *g* is the mirror image of *f* and yields an outcome worth 1 utile if ω_2 obtains and an outcome worth 0 utiles if ω_1 obtains. Option *h* is constant in value, yielding an outcome worth 0.4 utiles regardless whether ω_1 or ω_2 obtains. Figure 1 graphs the expected utilities for these three acts. Let $\mathcal{P}=P$ $0.25\leq P(\omega_1)\leq 0.75$. The surface of Bayes solutions is highlighted. The expected utility for options *f* and *g* each has the interval of values [0.25, 0.75], whereas h of course has constant expected utility of 0.4. From the choice set of these three options $\mathcal{A}=\{f, g, h\}$ the Γ-Maximin decision rule determines that h is (uniquely) best, assigning it a value of 0.4, whereas f and g each has a Γ-Maximin value of 0.25. By contrast, under *E*-admissibility, only the option h is *E-inadmissible* from the trio. Either of f or g is *E*-admissible. And, as no option is strictly preferred to any other by expectations with respect to $\mathcal{P}$, all three gambles are admissible under *Maximality*.

---

[15] Herzberger, "Ordinal Preferences and Rational Choice," *Econometrica*, XLI (1973): 187–234.

[16] Levi, "On Indeterminate Probabilities," and "Imprecise and Indeterminate Probabilities," in *Proceedings of the 1st International Symposium on Imprecise Proabilities and Their Applications* (1999).

[17] Seidenfeld, "Comment on 'Probability, Evidence and Judgement'," (by A.P. Dempster), in J.M. Bernardo, ed., *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting, 1983* (New York: North Holland, 1985), pp. 127–29.

[18] Sen, "Social Choice Theory: A Re-Examination," *Econometrica*, XLV (1977): 53–89.

We link this observation to the debate about the value of new information by considering a sequential decision problem in which the decision maker has the opportunity to postpone a terminal decision in order to learn the outcome of a *mixing* variable, a variable used to convexify the option space. Let the mixing variable α equal 1 or 0 as a *fair* coin lands Heads up or Tails up on a toss, so that $P(\alpha = 1) = P(\alpha = 0) = .5$. Assume, also, that α is independent of the states, Ω, over which the *pure* options are defined, so that each $P \in \mathcal{P}$, $P(\alpha, \omega) = .5P(\omega)$.

As a modification of Example 15, consider the mixed options m, and n, defined as follows:

$$m = \alpha f \oplus (1 - \alpha)g$$

$$n = \alpha g \oplus (1 - \alpha)f$$

Thus, m is the mixed act that uses the fair coin to bet on ω_1 if Heads (H) and to bet on ω_2 if Tails (T). Likewise, n is the dual mixed act that uses the same fair coin to bet on ω_2 if Heads and to bet on ω_1 if Tails. Note that m and n each carry a constant expected utility, .50. That is, these are options of determinate *risk*, despite the fact that uncertainty is with respect to the convex set of probabilities $\mathcal{P}$. Last, let the option *Status quo* denote no change in expected utility, with constant value 0.

The next two examples, sequential decision problems 1 and 2, are cast in extensive form. The three decision rules we discuss with these examples, each a generalization of Bayesian decision making when uncertainty is represented by a set of probability distributions rather than a single probability distribution, yields different results when extensive form decisions are changed to their normal form. Thus, in what follows we have a different situation than we saw in Examples 4 and 5. In those examples uncertainty is represented by a single probability distribution. Then, each of these three rules yields the same Bayesian analysis either in extensive form or in normal form.[19]

---

[19] See Seidenfeld, "When Normal and Extensive Form Decisions Differ" (*op. cit.*)

Sequential Decision Problem 1 The sequential decision problem, depicted in Figure 2, offers the decision maker one of three terminal options *m*– 40, *n*– 40, and *Status quo* The sequential option offers a 05 ***bonus*** to the decision maker to postpone this choice in order to learn the outcome of the mixing variable, α, and then to choose among these same three terminal options Of course, given $\alpha = 1$, act *m* is equivalent to act *f*, and act *n* is equivalent to act *g* Likewise, given $\alpha = 0$, act *m* is equivalent to act *g*, and act *n* is equivalent to act *f*

Let us reason through this decision problem using Γ-Maximin as the decision rule At either of the terminal choice nodes that might be reached under the sequential option, that is, after observing the coin flip, the decision problem reduces to a case formally analogous to the nonsequential problem of Example 15 That is, the constant option (worth 05) is uniquely Γ-Maximin admissible at either of these choice points Using this as the value assigned at the initial choice point to the sequential option, that is, by backward induction, we find that Γ-Maximin rejects the sequential option in favor of either *m*– 40 or *n*– 40, which are indifferent and worth 10 each Hence, Γ-Maximin mandates a negative value for the new information The Γ-Maximin

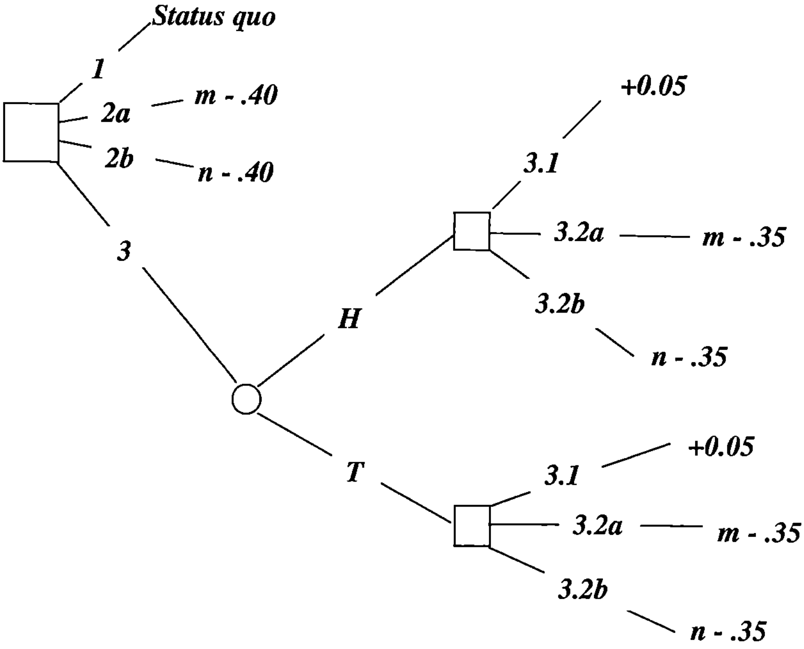


Figure 2 Sequential Decision Problem 1

decision maker is advised, in effect, to pay negative tuition pay not to learn!

By contrast, using *E*-admissibility as the decision rule, we find that the constant option is uniquely inadmissible at each of the choice points that might be reached under the sequential option Since, for example, *m*– 40 is inadmissible when *m*– 35 is available, by backward induction, the decision maker who uses *E*-admissibility will not assign a negative value to the new information available in this sequential decision problem

We are unsure what conclusion to draw about what *Maximality* recommends in this case At the two terminal choice points that might be reached under the sequential option, all three choices are Maximal Then, it is permissible for the decision maker who decides by Maximality to choose the constant (worth 05) if the sequential option is taken at the initial choice node But 05 is an inadmissible option at the initial choice node, since either *m*– 40 or *n*– 40 is strictly preferred to 05 Hence, it appears that in this sequential decision problem, Maximality does not require a decision maker to assign a non-negative value to potential cost-free information

We conclude this brief discussion of the value of new information by noting that even *E*-admissibility requires only that an admissible solution be Bayes in a *local*, but not in a *global* sense That is, in a sequential decision problem, where there are hypothetical decision nodes to consider that are not the initial choice node, *E*-admissibility does *not* require that there be one, common Bayes model for all combinations of *E*-admissible choices across these hypothetical decision nodes This is illustrated in Figure 3 by the following sequential decision problem, which is a variant of the previous one

Sequential Decision Problem 2 Here, at each of the two terminal choice nodes that might be reached under the sequential option, both options are *E*-admissible Then, it is permissible for the decision maker who decides by *E*-admissibility to choose the constant (worth 05) if the sequential option is taken at the initial choice node But 05 is an inadmissible option at the initial choice node, since *m*– 40 is strictly preferred to 05 Hence, it appears that in this sequential decision problem, *E*-admissibility does not require a decision maker to assign a non-negative value to potential cost-free information Of course, since, *E*-admissible options are always a subset of the options permissible by Maximality, to the extent that this phenomenon is a problem with *E*-admissibility as the decision rule, it is only more so of a problem for decision makers using Maximality

The examples of this section all involve countably additive probabilities and decisions about cost-free information Nevertheless, the

LIBRARY
UNIVERSITY OF CALCUTTA
ALIPORE CAMPUS

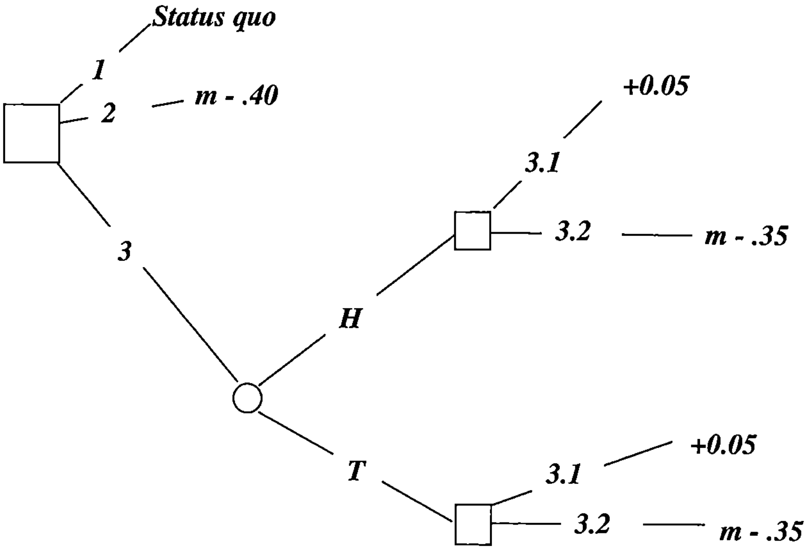


Figure 3 Sequential Decision Problem 2

theories of *E*-admissibility, maximality, and Γ-maximinity either allow or mandate assigning negative value to this cost-free information We are unable to see how these theories can justify assigning negative value to the cost-free information in these cases

VI IS IGNORANCE BLISS?

What would a world be like in which rational actors would pay not to receive sample information of certain kinds? It would take some effort to discover, for each candidate for such a service, just what information he or she would pay not to receive Suppose there were two companies that would offer this service, one for $0 50 and one for $0 25 To be effective, the consumer has to deal with both of them, and in fact with any other company also offering not to tell Indeed, the provider of this service could be most persuasive in guaranteeing to the customer that the data will not be provided, by not possessing it themselves This offers a new world of entrepreneurial activity of no apparent social value whatsoever

But are there instances in which a reasonable person would pay not to have certain information? Suppose you were offered the opportunity to learn the date and circumstances of your own death Would you pay not to have this information? For many of us, the answer is "yes " Such knowledge would change much about what it means to be human, and would make the experience of life rather different, in a profound way, from all of those who have gone before us To restructure the way we think about life, and its uncertain time of end, would be profoundly unsettling and costly, and hence it would be reasonable to pay not to have to undergo that process On the other hand, absent professional hitmen, none of us are likely to receive such an offer

The issue is more real, however, for those who have to decide whether to be tested for an incurable genetic disease such as Huntington's Disease This disease generally strikes in people of middle ages, with neurological and psychological symptoms, ultimately resulting in a period of helplessness followed by an early death There are limited treatments that can help to mitigate some of the symptoms, but no cure as yet Those who are genetic candidates for the disease have a 50% probability of having the gene Knowing that one is a candidate, is it reasonable not to have the test, if it were free? One reason not to have the test is the response of others to the results "people lose insurance and jobs "[20] But there are deeper reasons having to do with one's ability to cope and enjoy life that influence this decision We can not label either decision irrational, given all the implications

How can we defend the idea of refusing to pay not to see information and yet endorse a decision not to be tested for Huntington's disease by someone who has a parent who had the disease, and hence has a 50% probability of carrying the gene? The insurance issue is serious, since most life and health insurers ask about known preexisting conditions Hence someone who knows they have the gene could easily believe they would be denied coverage on that account However, one could imagine a strategy of buying insurance first, and then being tested Hence, although we take insurance to be an important practical problem, we do not take it as definitive The real reasons, we guess, lie deeper

EXAMPLE 16 (INHERITED DISEASE) There is a certain peace of mind associated with knowing that, aside from catastrophe, one has the ability to make plans and choices for a relatively long future There is also value to an individual of having the prospect of making choices and enjoying results in the future By having a short horizon placed on our ability to make choices and use our assets, we reduce the overall level of utility In the terminology of Sen,[21] the knowledge that one has the disease reduces one's opportunity-freedom In symbols, let ψ be the knowledge base sequence assuming that I will not learn the results of the test Let ψ' be the knowledge base sequence assuming that I will learn (at time t_1) whether or not I have the disease Let $X = 1$ if I have the disease and $X = 0$ if not Let $\Omega = \Theta \times \{0,1\}$, where the second coordinate gives the value of X Let $\theta \in \Theta$, and let δ be a decision that I could make at time $t > t_1$ regardless of the value of X It might be reasonable that

---

[20] Feldman, "One Family's Story A Quandry over Genetic Testing," *Houston Chronicle* (2003), http //www johnworldpeace com/e030112e htm

[21] Sen, "Markets and Freedoms Achievements and Limitations of the Market Mechanism in Promoting Individual Freedoms," *Oxford Economic Papers*, XLV (1993) 519–41, on p 522

$$(19) \qquad U((\theta,1),\psi_t',\delta) < U((\theta,1),\psi_t,\delta) < U((\theta,0),\psi_t,\delta) < U((\theta,0),\psi_t',\delta)$$

for all such θ and δ That is, knowing that I have the disease makes everything worth less than merely supposing that I have the disease, which in turn makes everything worth less than supposing that I do not have the disease, and knowing that I do not have the disease makes everything worth the most If I am not going to learn the results of the test, the expected utility now of choosing δ at time t is

$$(20)$$
$$E[U((\theta,X)\psi_t,\delta)] = 0.5E[U((\theta,0),\psi_t,\delta) \mid X = 0] + 0.5E[U((\theta,1),\psi_t,\delta) \mid X = 1]$$

On the other hand, if I am going to learn X in the future, the expected utility now of choosing δ at time t

$$(21) \qquad 0.5E[U((\theta,0),\psi_t',\delta) \mid X = 0] + 0.5E[U((\theta,1),\psi_t',\delta) \mid X = 1]$$

If the first inequality in (19) is satisfied by a much wider amount than the third inequality and is so satisfied uniformly in θ and δ, then the expected utility in (21) will be smaller than the one in (20) for all δ, hence one would prefer not to learn X

In this connection, it is interesting to note that this is the kind of information the poet Thomas Gray (1747) has in mind The quote at the start of this paper comes from a poem, "Ode on a Distant Prospect of Eton College " The first nine stanzas are about the careless optimism of the students at Eton, and the troubles they are likely to face in life The concluding tenth stanza reads

To each his sufferings all are men
Condemn'd alike to groan?
The tender for another's pain,
Th' unfeeling for his own
Yet, ah! why should they know their fate,
Since sorrow never comes too late
And happiness too swiftly flies?
Thought would destroy their Paradise
No more, where ignorance is bliss
'Tis folly to be wise

So the question remains of whether it is reasonable to impose the requirement on a theory of rational decision making that it not require or permit paying not to see cost-free data If it is, the only such theory known to us is Bayesian decision theory with a single countably-additive proper prior Each of the weakenings of this theory-allowing finitely additive priors, or improper ones, or allowing sets of probabilities with any one of the three decision rules studied here violates this principle

JOSEPH B KADANE
MARK SCHERVISH
TEDDY SEIDENFELD

Carnegie Mellon University

APPENDIX

A 1 Conditioning and Learning In this appendix we discuss two related matters for understanding how we use conditional probability in connection with changes in evidence

(i) what conditional probability means in connection with changing bodies of evidence, and
(ii) what it is that is *given* in a conditional probability

(i) The first issue about conditional probability arises both with randomized (so-called "mixed") options and in sequential decisions as we present those in section v We use coin flips that are stipulated to be independent of the other events of interest That is, for each joint probability distribution we consider, the conditional probability for the event of interest given the coin flip equals its unconditional probability In each example, we argue that, with respect to each probability distribution, given a result of the coin flip, the available options are valued the same as they are valued marginally under that distribution Such an argument is an implicit use of conditional probability There are at least three distinct interpretations of conditional probability, and this appendix attempts to identify which interpretation we have used in this paper

Let A and B be events in the algebra over which probability is defined, and let I_A and I_B be their indicator functions[22] The conditional probability $P(A|B)$, as it relates to degrees of belief and decision making, has been interpreted in at least the following three ways

---

[22] What we say here about conditional probability is not restricted to indicator random variables, which we use for simplicity of exposition only

(1) Conditional probability and *called-off preferences* In the theories of both DeFinetti and Savage, where probability is grounded on unconditional preferences, conditional probability is reduced to called-off preferences

(a) Specifically, in DeFinetti, $P(A|B)$ is that prevision p the decision maker offers in order to make the following bet "fair" for all values of α $I_B\alpha(I_A - p)$ The bet is "called-off" because there is neither a loss nor a gain if event B fails to occur DeFinetti shows that, in order for your called-off previsions to be coherent, that is, immune to a "Book," they must be the values of a (finitely additive) conditional probability function

(b) In Savage, conditional probability is identified with the decision maker's preferences over pairs of acts that are modified to agree with each other on states that comprise B^c, when B fails Thus, these pairs of acts conform to the requirements of being "called off" in the sense that, for states comprising B^c, the decision maker of Savage's theory receives the same regardless which "called-off" act is chosen

It is important to note that in each of these two theories there is no relation between conditional probability and updating beliefs by new evidence That is, neither of these theories goes beyond the statics of preference at a time Neither of these theories offers a dynamic interpretation of conditional probability that makes it into a learning rule

(2) Conditional probability and *temporal updating* It is a commonplace interpretation of Bayesian theory to offer the (current) conditional probability $P(A|B)$ as a "posterior" probability for what the investigator's degree of belief will be in A on the condition that he/she learns that B obtains Some attempts have been made to extend deFinetti's static "Book" argument to cover such cases, for example, to make conditional probability into a dynamical, learning rule See, for example, Levi,[23] for sound reasoning as to why this is an overreaching of the "Book" argument Levi refers to the dynamic interpretation as *temporal credal conditionalization*

(3) Conditional probability and *hypothetical reasoning* Yet a third sense of conditional probability, the sense used in this paper, is to understand $P(A|B)$ as the decision maker's current, hypothetical degrees of belief in A on the condition that he/she were to ac-

---

[23] Levi, "The Demons of Decision," *The Monist*, LXX (1987) 193–211

> cept B as true. This sense of conditional probability relates to the agent's current hypothetical preference were he/she to accept B as true. In this sense, the agent reasons from a hypothetically augmented body of knowledge that includes B as certain. It differs from the first sense of $P(\cdot \mid B)$ in that the space of serious possibility is changed by hypothesizing B as a certainty, equating B^c with the impossible event. It differs from the second sense of conditional probability in that the decision maker does not predict how he will update his/her degrees of belief at some later time, when B might be accepted as true. (In Levi, this is referred to as *confirmational conditionalization*.)

In short, these three senses of conditional probability $P(\cdot \mid B)$ differ in their consequences for decision making by addressing, respectively:

(1) current, unconditional called-off preferences, called off if B^c obtains,
(2) future, unconditional preferences, upon learning that B obtains,
(3) current, hypothetical preferences obtained by positing that B is true.

Throughout this paper, we use conditional probability to relate to decision making in the third sense, and we use the formalism of sequential decision making to highlight the difference between the first and third senses of conditional probability. We do not use (or advocate) the second sense of conditional probability.

(ii) The second matter involving our use of conditional probability in this paper is made salient in Example 12, the "Taxi Driver" problem. There is an important distinction between conditioning on who-done-it and adding the information to one's knowledge base. Think of D as a random variable taking the value i if the ith character in the mystery committed the crime. Conditioning on $D(\omega) = i$ for $i = 1, \ldots, \ell$ does not affect my expected utility for any of the decisions that I contemplate choosing in this decision problem. On the other hand, augmenting my current knowledge base either with the information that I know $D(\omega) = i$ or with the information that I will know $D(\omega) = i$ before getting to watch the play, has a pronounced effect on the value of going to the play. That knowledge affects my choices in the same, negative way regardless of the value of D that I (will) learn.

Thus, in our use of conditional probability, we need to distinguish conditioning from learning, and in two ways. From the first issue (i), we use conditional probability to model hypothetical changes in knowledge, rather than requiring conditional probability to be a learning rule. This takes us beyond merely the use of conditional probability to model called-off wagers. But it leaves us short of mandating a single dynamics by which rational agents learn.

From the second issue (ii), we must attend to the event on which we condition. We recognize that it is *not* generally the same to condition on E occurring, as it is to condition on the more inclusive event that you learned that E occurred by some particular method. The distinction between conditioning and learning has important consequences in the Taxi-Driver Example 12 and in the Inherited Disease Example 16. We believe that the distinction does not matter in the other examples discussed in this paper.

RELATIVE EXPECTATION THEORY*

The St Petersburg game and its nearby neighbors present serious problems for decision theory[1] The St Petersburg game invokes an unbounded utility function to produce an infinite expectation for playing the game The problem is usually presented as a clash between decision theory and intuition most people are not prepared to pay a large finite sum to buy into this game, yet this is precisely what decision theory suggests we ought to do But there is another problem associated with the St Petersburg game This second problem is that standard decision theory counsels us to be indifferent between any two acts that have infinite expected utility So, for example, consider the decision problem of whether to play the St Petersburg game or a game where every payoff is $1 higher Let us call this second game the *Petrograd game* (it is the same as St Petersburg but with a bit of twentieth-century inflation) Standard decision theory tells us to be indifferent between these two options

It might be argued that any intuition that the Petrograd game is better than the St Petersburg game is a result of misguided and naïve intuitions about infinity[2] But this argument against the intuition in question is misguided The Petrograd game is clearly better than the St Petersburg game And what is more, there is no confusion about infinity involved in thinking this When the series of coin tosses comes to an end (and it comes to an end with probability 1), no matter how many tails precede the first head, the payoff for the Petrograd game is one dollar higher than the St Petersburg game Whatever the

---

*Thanks to Alan Hájek, Adam La Caze, Aidan Lyon, Tony Martin, Martin Rechenauer, Katie Steele, and Juliana Weingaertner I am also grateful to audiences at the 2005 Stradbroke Island Philosophy and Linguistics Cogitamus, the 2005 Australasian Association of Philosophy (New Zealand) conference at the University of Otago in Dunedin, the Philosophy Department at the University of California, Los Angeles, and the Philosophy Department at the Ludwig-Maximilians University in Munich

[1] Recall that the St Petersburg game is a game where a fair coin is tossed repeatedly, if necessary, until the first head appears The game is then over The payoff for this game is $2 if the head appears on the first toss, $4 for heads on the second toss, $8 on the third, and so on See Robert Martin, "The St Petersburg Paradox," in Edward Zalta, ed , *The Stanford Encyclopedia of Philosophy (Fall 2001 Edition)*, URL = <http // plato stanford edu/archives/fall2001/entries/paradox-stpetersburg/>

[2] These are the same intuitions that incorrectly suggest that $1 + \infty > \infty$, or so the story goes

0022-362X/08/0501/37–44

outcome, you are better off playing the Petrograd game. Infinity has nothing to do with it. Indeed, a straightforward application of dominance reasoning backs up this line of reasoning.[3] Standard decision theory (in the form of expected utility maximization) cannot deliver this verdict, but so much the worse for standard decision theory. Dominance reasoning gets it right, and this is significant.

I have argued elsewhere[4] that there are a number of cases like this, where dominance reasoning delivers the correct result and expected utility maximization is either silent or indifferent. In light of such cases, it would seem that we have a kind of unsettling pluralism about decision rules on our hands: expected utility maximization and dominance reasoning are both required—neither will do on its own. So here is the first challenge: provide a generalization of decision theory that will unify the rule of expected utility maximization and the rule of dominance.

Next consider a variation on the Petrograd game. Like the Petrograd, this game has payoffs $1 higher than the corresponding payoffs of the St. Petersburg game—except for one. The exception is a payoff for some very low probability state and this is $1 *less* than the corresponding St. Petersburg payoff. Call this game the *Leningrad game*. Here expected utility theory suggests that we ought to be indifferent between the Leningrad game and the St. Petersburg game; dominance reasoning is not applicable and so is silent. But there is a very strong intuition that the Leningrad game is better than the St. Petersburg game. After all, the Leningrad game almost dominates the St. Petersburg game, and the probability of finding oneself in the nondominant state is, by construction, very low. So here is the second challenge: either find a decision rule that supports this intuition or explain away the intuition. But note how hard the latter will be. Any argument that explains away the intuition must not also explain away the intuition that the Petrograd game is better than the St. Petersburg game, for the latter intuition is certainly correct. So the real challenge here is to provide a decision-theoretic validation of the intuition that the Leningrad game is preferable to the St. Petersburg game.

In this paper I will present new rules of decision that meet both these challenges. That is, I will provide a generalization of decision theory that unifies expected utility theory and dominance reasoning Indeed, the theory I advance is a generalization in the sense that it is a conservative extension of both dominance reasoning and finite expected utility theory Moreover, I will show how the new theory is also able to deal with problem cases such as choosing between the Leningrad game and the St Petersburg game

---

[3] Recall that dominance reasoning suggests that one ought to choose act A_1 over act A_2 if in every state the utilities associated with A_1 are never less than the corresponding utilities for A_2, and in at least one state the utility of A_1 is higher than the corresponding utility for A_2. This rule is only applicable when the states are independent of the acts.

[4] Colyvan, "No Expectations," *Mind*, cxv (2006): 695–702.

I INTRODUCING RELATIVE EXPECTED UTILITY

Assume throughout that we have a decision problem where the states are independent of the acts We define the *relative expected utility* of act A_k over A_l as

$$\mathrm{REU}(A_k,A_l) = \sum_{i=1}^{n} p_i(u_{ki} - u_{li})$$

where p_i is the probability associated with state S_i and u_{ji} are the utilities of the outcome resulting from act A_j in state S_i For infinite state spaces we define relative expected utility similarly

$$\mathrm{REU}(A_k,A_l) = \sum_{i=1}^{\infty} p_i(u_{ki} - u_{li})$$

where the right-hand side converges or diverges to infinity The decision rule is then as follows

> REU Decision Rules Choose act A_k over act A_l iff $\mathrm{REU}(A_k,A_l) > 0$ If $\mathrm{REU}(A_k,A_l) = 0$ an agent should be indifferent between the two acts in question

The resulting theory I call *relative expectation theory* The basic idea here is clear Relative expected utility is the expected gain or loss in utility between the two acts in question In other words, it is the expected relative advantage in choosing one act over another[5] It is also worth noting that there is a sense in which the expected utility can be thought of as a special case of relative expected utility the expected utility of an act is the relative expected utility of that act over an act with zero payoffs for every state (the null act)[6] On the other hand, it might be argued that relative expectation theory is nothing new, economists may well have had something like this in mind when considering expected profits and losses, for example After all, since utilities are random variables, their differences are also random variables, and relative expectation theory just considers the expectation of these [7] While there is a sense in which this is right, it turns out that the advice relative expectation theory gives in the problem cases I opened with is quite different from that given by standard decision theory Standard decision theory counsels us to calculate the expected utility of each act and choose the act that has the greatest expected utility (if there is such an act) Relative expectation theory gives no such advice In fact, relative expectation theory does not even require the calculation of expectations for individual acts In the next section I will have more to say about the relationship between standard decision theory and relative expectation theory As we shall see, the two theories are closely related but the advice they give is quite different I will thus continue to speak of relative expectation theory as a distinct theory of decision So much for what relative expectation theory is, let me now say a little about what it is good for

---

[5] The basic idea of considering the expectations of differences was advanced independently in Alan Hájek, "In Memory of Richard Jeffrey Some Reminiscences and Some Reflections on *The Logic of Decision*," *Philosophy of Science*, LXXIII (2006) 947–58

[6] We need to be careful though For the standard axioms for utility theory, the von Neumann-Morgenstern axioms, do not allow us to make sense of a distinguished

It is straightforward to show that the above decision rules support the intuition to choose the Petrograd game over the St Petersburg game (since the relative expected utility of the Petrograd game over the St Petersburg game is $\Sigma_{i=1}^{\infty} 1/2^i=1$) It is also straightforward to show that so long as the nondominant utility in the Leningrad game is associated with a state with probability less than a half, we should choose the Leningrad game over the St Petersburg game It should also be intuitively clear that the above decision rules agree with dominance reasoning, whenever the latter applies, and agree with expected utility theory in at least cases with finite state spaces That is, relative expectation theory is a conservative extension of both the rule of dominance and finite expected utility theory (I will back up this last claim shortly)

Relative expected utility theory clearly has a lot going for it Indeed, it would seem to be a good candidate for the unification of dominance reasoning and expected utility theory which we seek It re-

---

zero See John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, 1944) Like temperature scales (and unlike distance scales), the zero of utility theory is purely conventional Thanks to Juliana Weingaertner for reminding me of this point

[7] Thanks to an anonymous referee for this suggestion

mains to demonstrate this though I do this in the next section by proving a number of theorems about relative expected utility

II SOME ELEMENTARY THEOREMS

First a couple of theorems that shows that our decision rules are well defined [8] That is, I show that the above rules do not give us contradictory advice

Theorem 1 If $\mathrm{REU}(A_1,A_2) > 0$, $\mathrm{REU}(A_2,A_1) < 0$

Proof Without loss of generality, assume that the state space is infinite $\mathrm{REU}(A_1,A_2) = \sum_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) > 0$, and since this series is defined it either converges or diverges to positive infinity [9] In either case $\mathrm{REU}(A_2,A_1) = \sum_{i=1}^{\infty} p_i(u_{2i} - u_{1i}) = -\mathrm{REU}(A_1,A_2) < 0$

A similar proof establishes a symmetry result for when the relative expected utility is zero

Theorem 2 If $\mathrm{REU}(A_1,A_2) = 0$, $\mathrm{REU}(A_2,A_1) = 0$

Having established that our new decision rules are well defined, I now prove a few results that establish the relationship between relative expectation theory and our earlier decision rules

Theorem 3 If $\mathrm{EU}(A_1) > \mathrm{EU}(A_2)$, $\mathrm{REU}(A_1,A_2) > 0$

Proof Again without loss of generality assume that the state space is infinite There are four cases to consider (i) $\mathrm{EU}(A_1)$ and $\mathrm{EU}(A_2)$ both converge, (ii) $\mathrm{EU}(A_1)$ diverges to infinity and $\mathrm{EU}(A_2)$ converges, (iii) $\mathrm{EU}(A_1)$ converges and $\mathrm{EU}(A_2)$ diverges to negative infinity, and (iv) $\mathrm{EU}(A_1)$ diverges to positive infinity and $\mathrm{EU}(A_2)$ diverges to negative infinity Consider case (i) first Since $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i} > \Sigma_{i=1}^{\infty} p_i u_{2i} = \mathrm{EU}(A2)$, $\Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$ In case (ii), we have $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i}$ is infinite and $\Sigma_{i=1}^{\infty} p_i u_{2i} = \mathrm{EU}(A_2)$ is finite, so $\mathrm{EU}(A_1) - \mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{1i} - \Sigma_{i=1}^{\infty} p_i u_{2i} = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$ In case (iii), we have $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i}$ is finite and $\mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{2i}$ is negatively infinite, so $\mathrm{EU}(A_1) - \mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{1i} - \Sigma_{i=1}^{\infty} p_i u_{2i} = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$ In case (iv), we have $\mathrm{EU}(A_1) = \Sigma_{i=1}^{\infty} p_i u_{1i}$ is positive infinite and $\Sigma_{i=1}^{\infty} p_i u_{2i} = \mathrm{EU}(A_2)$ is negative infinite, so $\mathrm{EU}(A_1) - \mathrm{EU}(A_2) = \Sigma_{i=1}^{\infty} p_i u_{1i} - \Sigma_{i=1}^{\infty} p_i u_{2i} = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = \mathrm{REU}(A_1,A_2) > 0$

And there is a similar proof that establishes the corresponding case for identity

---

[8] For all these theorems I continue to assume that the states are independent of the acts I will also make use of some results about well-behaved divergent series See G H Hardy, *Divergent Series* (New York Oxford, 1949) for more on this topic

[9] In this and other proofs, I adopt the common convention of treating a series that diverges to infinity as being well defined and having infinity as its limit

Theorem 4 If $\mathrm{EU}(A_1) = \mathrm{EU}(A_2)$ and is finite, then $\mathrm{REU}(A_1,A_2) = \mathrm{REU}(A_2,A_1) = 0$

These two results tell us that relative expectation theory agrees with expected utility theory in finite cases (and whenever the states are independent of the acts) It is important to note that theorem 3 holds only in the direction stated Indeed, it is the fact that relative expectation theory is more discriminating than expected utility theory that allows us to use relative expectation theory to justify choosing the Petrograd game over the St Petersburg game Inequalities in relative expected utilities do not imply inequalities in expected utilities and we do not require the expectation series to converge in order to make meaningful comparisons of actions

Next a theorem relating relative expected utility to dominance

Theorem 5 If A_1 dominates A_2, $\mathrm{REU}(A_1,A_2) > 0$

Proof Again without loss of generality assume that the state space is infinite Since A_1 dominates A_2, each u_{1i} is greater than or equal to each u_{2i} and at least one u_{1i} is strictly greater than a corresponding u_{2i} But this implies that $\mathrm{REU}(A_1,A_2) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) > 0$

This result tells us that relative expectation theory agrees with dominance reasoning

Next a couple of transitivity results

Theorem 6 If $\mathrm{REU}(A_1,A_2) > 0$ and $\mathrm{REU}(A_2,A_3) > 0$, then $\mathrm{REU}(A_1,A_3) > 0$

Proof Once again, without loss of generality, assume that the state space is infinite $\mathrm{REU}(A_1,A_2) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) > 0$ and $\mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) > 0$ It follows that $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) > 0$ But since the series in question either converge or diverge to positive infinity, $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) + \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{3i}) = \mathrm{REU}(A_1,A_3) > 0$

Theorem 7 If $\mathrm{REU}(A_1,A_2) = 0$ and $\mathrm{REU}(A_2,A_3) = 0$, then $\mathrm{REU}(A_1,A_3) = 0$

Proof Once again, without loss of generality, assume that the state space is infinite $\mathrm{REU}(A_1,A_2) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) = 0$ and $\mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) = 0$ It follows that $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) = 0$ But $\mathrm{REU}(A_1,A_2) + \mathrm{REU}(A_2,A_3) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{2i}) + \Sigma_{i=1}^{\infty} p_i(u_{2i} - u_{3i}) = \Sigma_{i=1}^{\infty} p_i(u_{1i} - u_{3i}) = \mathrm{REU}(A_1,A_3) = 0$

Finally, the promised conservativeness result

Theorem 8 Relative expected utility theory is a conservative extension of both finite expected utility theory and dominance reasoning

Proof This follows from theorems 3, 4, and 5

III DISCUSSION

Theorem 8 is the crucial one This shows that relative expectation theory is the generalization we seek Moreover, relative expectation theory provides a satisfying account of the near-dominance cases such as the Leningrad game versus the St Petersburg game And it does this in a unified way There is no special pleading for problem cases [10]

There are, however, some limitations to this approach that deserve commenting on The first is the assumption that the states are independent of the acts I relied on this assumption in the very definition of relative expected utility, when I required that the probabilities for each state do not depend on the acts Dispensing with this assumption and thus generalizing the notion of relative expected utility would free things up, but it is not obvious how to do this In any case, as things stand the definition of relative expected utility and thus the proofs of the theorems relied on the independence of states and acts, so we must live with this assumption for the time being at least [11]

The second restriction is that the new theory can only deal with pair-wise comparisons of acts Indeed, the very definition of relative expected utility only makes sense of pair-wise comparisons In light of theorems 6 and 7, though, this is not such a serious limitation For theorems 6 and 7 show us that we can make sense of decision problems with more than 2 acts—it is just that we need to compare the acts two-by-two Theorems 6 and 7 tell us, in effect, that we can produce an ordering of the acts, and since any such ordering of finitely many acts will have a (perhaps nonunique) maximal act, we can solve the decision problem despite the limitation of using only pair-wise comparisons [12]

---

[10] Although I will not pursue the details here, relative expectation theory also deals with a class of problems where expected utility theory is silent I have in mind here the Pasadena game versus the Altadena game See Harris Nover and Alan Hájek, "Vexing Expectations," *Mind*, CXIII (2004) 237–49, Hájek and Nover, "Perplexing Expectations," *Mind*, CXV (2006) 703–20, and Colyvan, "No Expectations " That this is so follows straightforwardly from theorem 5 above

[11] This issue also touches on the treatment of Newcomb's paradox, see Robert Nozick, "Newcomb's Problem and Two Principles of Choice," in Nicholas Rescher, ed , *Essays in Honour of Carl G Hempel* (Boston Reidel, 1969), pp 114–46 Since Newcomb's problem places dominance reasoning and expected utility theory in conflict with one another, it would suggest that dropping the independence condition will be no easy matter For example, if we could prove analogues of theorems 3 and 5 without the independence assumption, Newcomb's problem would present itself as a serious problem for relative expectation theory Perhaps what is required is a causal version of relative expectation theory, see James Joyce, *The Foundations of Causal Decision Theory* (New York Cambridge, 1999)

[12] It is important to note that this is only the case when all the relative expected utilities are defined There will be cases when they are not Not surprisingly, relative

These limitations suggest that there is further work to be done here, but still relative expectation theory is genuinely useful in solving some persistent and troubling problems in infinite decision theory

MARK COLYVAN

University of Sydney

---

expectation theory will not be able to solve these problems (Consider, for example, the case of play or not to play the Pasadena game Since the Pasadena game does not have an expectation, there is no relative expectation for playing the Pasadena game over not playing it See Nover and Hájek, "Vexing Expectations," Hájek and Nover, "Perplexing Expectations," and Colyvan, "No Expectations ")

NOTES AND NEWS

The editors of *The Journal of Philosophy* take note with sorrow of the death of Henry E. Kyburg on October 30, 2007. At the time of his death, Kyburg was Gideon Burbank Professor of Moral Philosophy and Professor of Computer Science at the University of Rochester, New York, and Pace Eminent Scholar at The Institute of Human and Machine Cognition, Pensacola, Florida.

Kyburg received his Ph.D. in Philosophy from Columbia University in 1955 under the supervision of Ernest Nagel. His doctoral dissertation was *Probability and Induction in the Cambridge School*. Subsequently he was an Assistant Professor in Mathematics at Wesleyan University (1958–61), Research Associate, Rockefeller Institute (1961–62), Associate Professor of Mathematics and Philosophy, University of Denver (1962–63), Associate Professor of Philosophy, Wayne State University (1963–65), Professor of Philosophy (1965–2007), Burbank Professor of Moral and Intellectual Philosophy (1982–2007), and Professor of Computer Science (1986–2007) at the University of Rochester. Kyburg was chair of the Philosophy Department at Rochester from 1969 to 1982. He became a Fellow of the American Association for the Advancement of Science in 1982, a Fellow of the American Academy of Arts and Sciences in 1995, and a Fellow of the American Association for Artificial Intelligence in 2002. In 1982, he was awarded the Butler Medal for Philosophy in Silver by Columbia University.

Kyburg's distinctive ideas concerning induction and probability were elaborated in *Probability and the Logic of Rational Belief* (1961) and *The Logical Foundations of Statistical Inference* (1979). These books set the stage for the topics and issues Kyburg addressed in his subsequent writings. Some of these later works (such as *Uncertain Inference*, coauthored with Choh Man Teng (2001), and the anthology *Knowledge Representation and Defeasible Reasoning*, coedited with Greg Carlson and Ronald Loui (1990)), were focused on relating his ideas to topics of interest in computer science. *Theory and Measurement* (1983), *Science and Reason* (1990), and the collection of papers *Epistemology and Inference* (1983) embed his ideas on probability and induction in a sophisticated account of measurement, scientific theories and topics in epistemology. The collection of essays, *Probability Is the Very Guide in Life*, coedited with Mariam Thalos (2002), emphasizes the practical applications of Kyburg's conception of probability. Kyburg was a coeditor of several other works including a widely read and referenced anthology, *Studies in Subjective Probability*, that he coedited with Howard E. Smokler. This volume includes Kyburg's translation into English of Bruno de Finetti's *La Prevision*. This was the first such translation of this canonical testament of the foundation of subjective or personal probability.

Several leading ideas guided the view that Kyburg constructed: (1) Judgments of belief probability ought to be grounded in information accepted by the inquirer about frequencies of characteristic traits in sets of some kind. (2) In agreement with De Finetti, Kyburg maintained that chance or statistical probability that other authors like Peirce thought should ground judgments of belief probability is metaphysical nonsense. Unlike De Finetti, however, he thought it could be replaced by information about frequencies so that, again unlike De Finetti, Savage, and other personalists, he held that probability judgment could always be grounded in facts. (3) However, the judgments of belief probability grounded in this way are, in general, interval-valued. (4) Information about frequencies in one reference class often conflict with information in another. Precise information about frequencies in a broader class takes precedence over less precise information in a narrower reference class. This allows for "updating" judgments of belief probability that are relatively indeterminate initially so that the new probabilities are represented by narrower intervals. Such updating will, in general, violate Bayesian prescriptions for updating probabilities on the basis of new data in a manner reminiscent of the proposals by R.A. Fisher for assigning probabilities on the basis of fiducial arguments. (5) When the lower probability of the interval assigned a hypothesis about a frequency is above an appropriate threshold, the inquirer will be entitled to accept the hypothesis and use it ground further judgments of belief probability. Kyburg noted that the set of propositions whose lower probabilities are above a given threshold will, in general, fail to be closed under deduction. He illustrated this feature of his view by considering a fair lottery with *n* tickets (*n* large) and pointed out that the high probability rule leads to predicting that "ticket *i* will not win" for each *i*. This example became the famous lottery paradox.

Kyburg the scholar was also an important teacher. His students at the University of Rochester included Prasanta Bandyopadhyay, William Harper, Gary Merrill, Ron Loui, Abhaya Nayak, Choh Man Teng, Teddy Seidenfeld, Mariam Thalos, and Gregory Wheeler.

Kyburg, the scholar and teacher, was also a sailor, an ice hockey player and a farmer who together with his wife Sarah, raised Black Angus cattle on their farm in Lyons, NY, where they were gracious hosts entertaining friends and colleagues as well as students in a way that guests would always find memorable.

Kyburg's work was characterized by originality, independence, great learning, a refusal to sacrifice accuracy and precision for the sake of an easy accessibility. He remains a model of integrity to all who knew him. And although he was clearly passionately devoted to his wife, his children, his farm and his work, Henry was a gentle man with a quiet sense of humor who reminded all of us of the importance of exercising critical distance in philosophical reflection.

His presence will be missed by his many friends and colleagues.

MEMO TO AUTHORS

The Journal of Philosophy is published monthly, in issues that (ordinarily) contain only 56 pages of copy, this limits the length of papers the editors can normally publish to no more than 7,500 words, or 18 *Journal* pages. A more precise way of determining length is by characters (including spaces): 2,850 characters per *Journal* page, or 51,300 characters per *Journal* article of 18 pages.

The editors request that three copies (or an electronic version) of all manuscripts submitted be prepared according to the following guidelines:

(1) All material, including the quotations and footnotes, must be typed, double-spaced, and one-sided, with ample margins left and right, top and bottom.
(2) No bibliographies will be published, bibliographical material should be put into footnotes.
(3) Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include: name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include: publisher, U.S. city of publication, and date. Footnotes should be gathered at the end of the article.
(4) Quotations should be checked by the author, capitalization, italics, spelling, and the like should match the original.
(5) If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked. Standard logic symbols are available, and letters of the Greek and script alphabet.
(6) No underlining, please limit use of italics to technical terms (when they are introduced) and to variables.
(7) Minimize use of contractions.

The editors are willing to read and evaluate a manuscript that does not meet these guidelines, provided the author agrees to meet them before submitting a final manuscript. The editors may ask for revision of excessively long manuscripts, and may require reformatting of accepted manuscripts. They also request an electronic version with the final hard copy.

The Journal of Philosophy does not publish papers that are primarily historical or expository, the editors urge that such manuscripts be submitted elsewhere. While we do not publish manuscripts that mainly comment on articles published elsewhere, we do have a "Comments and Criticism" section in which short responses to *Journal* articles may be published.

Manuscripts may be sent to John Smylie, Managing Editor, *The Journal of Philosophy*, 709 Philosophy Hall, 1150 Amsterdam Avenue, Mail Code 4972, New York, NY 10027, Email jys3@columbia.edu

NEW BOOKS ANTHOLOGIES

ADAMS, FREDERICK, and KENNETH AIZAWA *The Bounds of Cognition* Malden, MA Blackwell, 2007 xiii, 197 p Cloth $74 95

AHMED, ARIF *Saul Kripke Contemporary American Thinkers* New York Continuum, 2007 182 p Cloth $29 95

ALEXANDER, J MCKENZIE *The Structural Evolution of Morality* New York Cambridge, 2007 ix, 300 p Cloth $95 00

ALMÉR, ALEXANDER *Naturalizing Intentionality Inquiries into Realism and Relativism* Göteborg, Sweden University Press, 2007 x, 282 p NP

BAKER, LYNNE RUDDER *The Metaphysics of Everyday Life* New York Cambridge, 2007 xv, 253 p Cloth $85 00

BARSKY, ROBERT F *The Chomsky Effect A Radical Works Beyond the Ivory Tower* Cambridge MIT, 2007 xviii, 381 p Cloth $29 95

BATLIN, MARGARET P , et al *Drugs and Justice Seeking a Consistent, Coherent, Comprehensive Worldview* New York Oxford, 2007 xxviii, 279 p Paper $21 95

BEALS, COREY *Levinas and the Wisdom of Love* Waco, TX Baylor, 2007 xiii, 169 p Cloth $29 95

BENZONI, FRANCISCO J *Ecological Ethics and the Human Soul* Notre Dame University Press, 2007 vi, 261 p Cloth $35 00

BIRD, ALEXANDER *Nature's Metaphysics Laws and Properties* New York Oxford, 2007 ix, 231 p Cloth $49 00

BOWIE, ANDREW *Music, Philosophy, and Modernity* New York Cambridge, 2007 xiii, 428 p Cloth $99 00

BRECHER, BOB *Torture and the Ticking Bomb* Malden, MA Blackwell, 2007 x, 122 p Paper $18 00

BROADIE, SARAH *Aristotle and Beyond Essays on Metaphysics and Ethics* New York Cambridge, 2007 x, 344 p Cloth $85 00

BUECHNER, JEFF *Godel, Putnam, and Functionalism A New Reading of Representation and Reality* Cambridge MIT, 2007 344 p Cloth $38 00

CALCAGNO, ANTONIO *Badiou and Derrida* New York Continuum, 2007 136 p Cloth $110 00

CAREY, ROSALIND *Russell and Wittgenstein on the Nature of Judgment* New York Continuum, 2007 150 p Cloth $110 00

CARROLL, NOËL *The Philosophy of Motion Pictures* Malden, MA Blackwell, 2007 241 p Cloth $35 00

CARUS, A W *Carnap and Twentieth-Century Thought* New York Cambridge, 2007 xiv, 346 p Cloth $95 00

CASEY, EDWARD S *The World at a Glance* Bloomington Indiana UP, 2007 xv, 498 p Paper $34 95

CHAPLIN, TAMARA *Turning on the Mind French Philosophers on Television* Chicago University Press, 2007 xi, 334 p Cloth $62 50, paper $24 00

CHAKRAVARTTY, ANJAN *A Metaphysics for Scientific Realism* New York Cambridge, 2007 xvii, 251 p Cloth $85 00

CHATEAUBRIAND, OSWALDO *Logical Forms, Part II Logic, Language, and Knowledge* Rio de Janeiro Coleção Cle, 2005 521 p NP

CHEAH, PHENG *Inhuman Conditions On Cosmopolitanism and Human Rights* Cambridge Harvard, 2006 vi, 321 p Paper $24 95

CHIESA, LORENZO *Subjectivity and Otherness A Philosophical Reading of Lacan* Cambridge MIT, 2007 235 p Cloth $19 95

CHURCHLAND, PAUL *Neurophilosophy at Work* New York Cambridge, 2007 xi, 249 p Cloth $75 00, paper $24 99

CI, JIWEI *The Two Faces of Justice* Cambridge Harvard, 2006 viii, 252 p Cloth $39 95

CLARKE, D S *Some Pragmatist Themes* Lanham, MD Lexington, 2007 xii, 102 p Cloth $55 00

CLARKE, DESMOND M *Descartes A Biography* New York Cambridge, 2006 xi, 507 p Cloth $40 00

CLARK, W NORRIS *The Philosophical Approach to God A New Thomistic Perspective* New York Fordham, 2007 viii, 163 p Paper $20 00

CLAYTON, JOHN *Religions, Reasons, and Gods Essays in Cross-Cultural Philosophy of Religion* New York Cambridge, 2006 xix, 372 p Cloth $100 00

CODE, LORRAINE *Ecological Thinking The Politics of Epistemic Location* New York Oxford, 2006 xiii, 322 p Cloth $74 00, paper $35 00

COHEN, ELLIOT D *The New Rational Therapy Thinking Your Way to Serenity, Success, and Profound Happiness* Lanham, MD Rowman and Littlefield, 2007 xiii, 294 p Cloth $70 00, paper $24 95

COHN-SHERBOK, DAN *Fifty Key Jewish Thinkers* New York Routledge, 2007 xii, 207 p Paper $27 95

COLE, PHILLIP *The Myth of Evil Demonizing the Enemy* Westport, CT Praeger, 2006 viii, 256 p Cloth $49 95

COLLINGWOOD, R G *The Philosophy of Enchantment Studies in Folktale, Cultural Criticism, and Anthropology* New York Oxford, 2005 cxix, 380 p Cloth $95 00

COLLINS, JEFFREY R *The Allegiance of Thomas Hobbes* New York Oxford, 2005 xii, 313 p Cloth $99 00

COLLINS, SUSAN D *Aristotle and the Rediscovery of Citizenship* New York Cambridge, 2006 x, 193 p Cloth $70 00

CONDREN, CONAL *Argument and Authority in Early Modern England The Presupposition of Oaths and Offices* New York Cambridge, 2006 x, 399 p Cloth $95 00

CONSTABLE, MARIANNE *Just Silences The Limits and Possibilities of Modern Law* Princeton University Press, 2005 x, 207 p Cloth $29 95

COOK, J A *Global Government Under the U S Constitution* Lanham, MD University Press of America, 2007 x, 102 p Paper $19 95

COOK, J THOMAS *Spinoza's Ethics* New York Continuum, 2007 174 p Paper $14 95

COOK, JOHN W *The Undiscovered Wittgenstein The Twentieth Century's Most Misunderstood Philosopher* Amherst, NY Humanity, 2006 437 p Cloth $55 00

COOKE, ELIZABETH F *Peirce's Pragmatic Theory of Inquiry Fallibilism and Indeterminacy* New York Continuum, 2007 xii, 174 p Cloth $120 00

COOKE, MAEVE *Re-Presenting the Good Society* Cambridge MIT, 2006 264 p Cloth $35 00

COOKSEY, THOMAS L *Masterpieces of Philosophical Literature* Westport, CT Greenwood, 2006 244 p Cloth $49 95

COOPE, CHRISTOPHER MILES *Worth and Welfare in the Controversy over Abortion* New York Palgrave Macmillan, 2006 x, 350 p Cloth $68 00

COOPER, DAVID E *A Philosophy of Gardens* New York Oxford, 2006 viii, 173 p Cloth $35 00

COPP, DAVID *Morality in a Natural World Selected Essays in Metaethics* New York Cambridge, 2007 xiii, 361 p Cloth $85 00

CORLETT, J ANGELO *Interpreting Plato's Dialogues* Las Vegas Parmenides, 2005 xii, 137 p Cloth $28 00

COSTELLOE, TIMOTHY M *Aesthetics and Morals in the Philosophy of David Hume* New York Routledge, 2007 xvii, 138 p Cloth $135 00

COTTINGHAM, JOHN *The Spiritual Dimension Religion, Philosophy and Human Value* New York, Cambridge, 2005 xii, 186 p Cloth $70 00, paper $24 99

COVENTRY, ANGELA *Hume A Guide for the Perplexed* New York Continuum, 2007 xi, 178 p Paper $19 95, *Hume's Theory of Causation A Quasi-Realist Interpretation* New York Continuum, 2006 xi, 166 p Cloth $120 00

COX, CHANA B *Liberty God's Gift to Humanity* Lanham, MD Rowman & Littlefield, 2006 285 p Cloth $90 00, paper $34 95, *Reflections on the Logic of the Good* Lanham, MD Rowman & Littlefield, 2007 xvii, 197 p Cloth $65 00

CRAGG, KENNETH *God's Wrong Is Most of All* Brighton, UK Sussex Academic, 2006 x, 201 p Paper $29 95

CRAIG, DAVID M *John Ruskin and the Ethics of Consumption* Charlottesville Virginia UP, 2006 x, 422 p Cloth $60 00

CRIBB, ALAN *Health and the Good Society Setting Healthcare Ethics in Social Context* New York Oxford, 2005 xiv, 236 p Cloth $85 00

CRISP, ROGER *Reasons and the Good* New York Oxford, 2006 x, 178 p Cloth $45 00

CRITCHLEY, SIMON *Infinitely Demanding Ethics of Commitment, Politics of Resistance* New York Verso, 2007 viii, 168 p Cloth $26 95

CROCKETT, CLAYTON *Interstices of the Sublime Theology and Psychoanalytic Theory* New York Fordham, 2007 xii, 216 p Cloth $70 00, paper $26 00

CROUTER, RICHARD *Friedrich Schleiermacher Between Enlightenment and Romanticism* New York Cambridge, 2005 xi, 277 p Cloth $80 00

CROWE, BENJAMIN D *Heidegger's Religious Origins Destruction and Authenticity* Bloomington Indiana UP, 2006 xv, 320 p Cloth $65 00, paper $24 95

CROWTHER, PAUL *Defining Art, Creating the Canon Artistic Value in an Era of Doubt* New York Oxford, 2007 268 p Cloth $65 00

CUDD, ANN E *Analyzing Oppression* New York Oxford, 2006 xiii, 278 p Cloth $74 00, paper $35 00

CUNEO, TERENCE *The Normative Web An Argument for Moral Realism* New York Oxford, 2007 263 p Cloth $63 00

CUNNINGHAM, CONOR *Genealogy of Nihilism Philosophies of Nothing and the Difference of Theology* New York Routledge, 2002 xviii, 315 p Cloth $120 00, paper $39 95

CURRY, MARK *About Time Narrative, Fiction and the Philosophy of Time* Edinburgh, UK University Press, 2007 ix, 160 p Cloth $80 00

CURRY, PATRICK *Ecological Ethics An Introduction* Malden, MA Polity, 2005 x, 173 p Paper $22 95

DALLMAYR, FRED *In Search of the Good Life A Pedagogy for Troubled Times* Lexington Kentucky UP, 2007 xvi, 320 p Cloth $40 00, *Small Wonder Global Power and Its Discontents* New York Rowman and Littlefield, 2005 xi, 241 p Cloth $75 00, paper $27 95

DANIELS, NORMAN *Just Health Meeting Health Needs Fairly* New York Cambridge 2007 ix, 397 p Cloth, $80 00, paper $29 99

DANISH, ROBERT *Pragmatism, Democracy, and the Necessity of Rhetoric* Columbia South Carolina UP, 2007 xii, 190 p Cloth $39 95

DARDEN, LINDEY *Reasoning in Biological Discoveries Essays on Mechanisms, Inter-field Relations, and Anomaly Resolution* New York Cambridge, 2006 xxiv, 346 p Cloth $75 00

DARWALL, STEPHEN *The Second-Person Standpoint Morality, Respect, and Accountability* Cambridge Harvard, 2006 xii, 348 p Cloth $49 95

DASTON, LORRAINE, and PETER GALISON *Objectivity* New York Zone, 2007 501 p Cloth $38 95

DAVENPORT, JOHN *Will as Commitment and Resolve* New York Fordham, 2007 xxiv, 706 p Cloth $95 00

DAVIDSON, DONALD *The Essential Davidson* ERNIE LEPORE and KIRK LUDWIG, eds New York Oxford, 2006 282 p Paper $26 69

DAVIES, STEPHEN *Philosophical Perspectives on Art* New York Oxford, 2007 279 p Cloth $60 00, *The Philosophy of Art* Malden, MA Blackwell, 2006 xi, 239 p Paper $24 95

DAVIS, WAYNE A *Nondescriptive Meaning and Reference An Ideational Semantics* New York Oxford, 2005 viii, 450 p Cloth $99 00

DAVIS, WILLIAM ERIC *Peace and Prosperity in an Age of Incivility* Lanham, MD University Press of America, 2006 x, 323 p Paper $39 95

DAWSON, HANNAH *Locke, Language and Early-Modern Philosophy* New York Cambridge, 2007 361 p Cloth $90 00

DE GAYNESFORD, MAXIMILIAN *I The Meaning of the First Person Term.* New York Oxford, 2006 viv, 198 p Cloth $65 00

DEAKIN, MICHAEL A B *Hypatia of Alexandria Mathematician and Martyr* Amherst, NY Prometheus, 2007 231 p Cloth $28 00

DEAN, RICHARD *The Value of Humanity in Kant's Moral Theory* New York Oxford, 2006 x, 270 p Cloth $74 00

DEMBSKI, WILLIAM A *The Design Inference Eliminating Chance through Small Probabilities* New York Cambridge, 2006 xvii, 243 p Cloth $85 00, paper $24 99

DE SOUSA, RONALD *Why Think? Evolution and the Rational Mind* New York Oxford, 2007 187 p Cloth $25 00

D'ESPAGNAT, BERNARD *On Physics and Philosophy* Princeton University Press, 2006 ix, 503 p Cloth $35 00

DEVITT, MICHAEL *Ignorance of Language* New York Oxford, 2006 xiii, 304 p Cloth $55 00

DEWEESE, GARRETT J , and J P MORELAND *Philosophy Made Slightly Less Difficult A Beginner's Guide to Life's Big Questions* Downers Grove, IL InterVarsity, 2005 170 p Paper $15 00

DEWEY, JOHN *Essays in Experimental Logic* D MICAH HESTER AND ROBERT B TALIESSE, eds Carbondale Southern Illinois UP, 2007 xxxv, 303 p Cloth $55 00, DEWEY, JOHN *The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought* LARRY A HICKMAN, ed Carbondale Southern Illinois UP, 2007 xxxii, 246 p Cloth $55 00

DIENSTAG, JOSHUA FOA *Pessimism Philosophy, Ethic, Spirit* Princeton University Press, 2006 xix, 293 p Cloth $35 00

DIETHE, CAROL *Historical Dictionary of Nietzscheanism Second Edition* Lanham, MD Scarecrow, 2007 lvii, 358 p Cloth $90 00

DILWORTH, CRAIG *The Metaphysics of Science An Account of Modern Science in Terms of Principles, Laws, and Theories* Dordrecht, Netherlands Springer, 2007 xvi, 333 p Paper $69 95

DODD, JULIAN *Works of Music An Essay in Ontology* New York Oxford, 2007 xi, 286 p Cloth $65 00

DOLAR, MLADEN *A Voice and Nothing More* Cambridge MIT, 2006 216 p Paper $19 95

DOLEV, YUVAL *Time and Realism* Cambridge MIT, 2007 ix, 237 p Paper $30 00

DOMBROWSKI, DANIEL A *Rethinking the Ontological Argument A Neoclassical Theistic Response* New York Cambridge, 2006 vii, 172 p Cloth $70 00

DOWNING, CRYSTAL L *How Postmodernism Serves (My) Faith Questioning Truth in Language, Philosophy and Art* Downers Grove, IL InterVarsity, 2006 240 p Paper $18 00

DOYLE, JON *Extending Mechanics to Minds The Mechanical Foundations of Psychology and Economics* New York Cambridge, 2006 xxi, 453 p Cloth $95 00

DRESCHER, GARY L *Good and Real Demystifying Paradoxes from Physics to Ethics* Cambridge MIT, 2006 xiv, 347 p Cloth $35 00

DRIVER, JULIA *Ethics The Fundamentals* Malden, MA Blackwell, 2006 vi, 286 p Paper $24 95

DUMMETT, MICHAEL *Thought and Reality* New York Oxford, 2006 ix, 109 p Cloth $29 95

DURAN, JANE *Eight Women Philosophers Theory, Politics, and Feminism* Urbana Illinois UP, 2006 xiv, 308 p Cloth $65 00, paper $29 95

DUSEK, VAL *Philosophy of Technology An Introduction* Malden, MA Blackwell, 2006 v, 244 p Cloth $74 95, paper $29 95

DYSON, FREEMAN J *A Many-Colored Glass Reflections on the Place of Life in the Universe* Charlottesville Virginia UP, 2007 xi, 162 p Cloth $21 95

EAGLETON, TERRY *The Meaning of Life* New York Oxford, 2007 x, 187 p Cloth $19 95

EARNSHAW, STEVEN *Existentialism A Guide for the Perplexed* New York Continuum, 2006 vi, 186 p Cloth $89 95, paper $19 95

EGER, MARTIN *Science, Understanding, and Justice* ABNER SHIMONY, ed Chicago Open Court, 2006 xxix, 538 p Paper $46 95

ELBOURNE, PAUL D *Situations and Individuals* Cambridge MIT, 2005 xii, 234 p Paper $32 00

ELLIOTT, DENI *Ethics in the First Person A Guide to Teaching and Learning Practical Ethics* Lanham, MD Rowman and Littlefield, 2006 vii, 139 p Cloth $72 00, paper $24 95

ELSTER, JON *Explaining Social Behavior More Nuts and Bolts for the Social Sciences* New York Cambridge, 2007 xi, 484 p Cloth $27 99, paper $27 99

EKKAN, KAIBARA *The Philosophy of QI The Record of Great Doubts* New York Columbia, 2007 xi, 208 p Cloth $34 50

EMILSSON, EYJOLFUR KJALAR *Plotinus on Intellect* New York Oxford, 2007 232 p Cloth $65 00

EVANS, C STEPHEN *Kierkegaard on Faith and the Self Collected Essays* Waco, TX Baylor, 2006 xv, 385 p Cloth $49 95

EVRIGENIS, IOANNIS D *Fear of Enemies and Collective Action* New York Cambridge, 2007 xix, 232 p Cloth $85 00

FABRE, CECILE *Whose Body Is It Anyway? Justice and the Integrity of the Person* New York Oxford, 2006 xiv, 232 p Cloth $55 00

FALCON, ANDREA *Aristotle and the Since of Nature Unity without Uniformity* New York, Cambridge, 2005 xvii, 139 p Cloth $75 00

FAIRFIELD, PAUL *Public/Private* Lanham, MD Rowman and Littlefield, 2005 v, 153 p Cloth $60 00, paper $22 95

FEDER, ELLEN K *Family Bonds Genealogies of Race and Gender* New York Oxford, 2007 xi, 145 p Paper $29 95

FELLENZ, MARC R *The Moral Menagerie Philosophy and Animal Rights* Urbana Illinois UP, 2007 301 p Cloth $60 00, paper $25 00

FELT, JAMES W *Aims A Brief Metaphysics for Today* Notre Dame University Press, 2007 xxii, 148 p Paper $20 00

FERGUSON, KENNAN *William James Politics in the Pluriverse* Lanham, MD Rowman and Littlefield, 2007 xxvii, 111 p Cloth $70 00, paper $24 95

FETZER, JAMES H *Render unto Darwin Philosophical Aspects of the Christian Right's Crusade against Science* Chicago Open Court, 2007 xix, 201 p Paper $24 95

FIENGO, ROBERT, and ROBERT MAY *De Lingua Belief* Cambridge MIT, 2007 viii, 179 p Cloth $32 00

FINE, KIT *Modality and Tense Philosophical Papers* New York Oxford, 2005 387 p Cloth $99 00, paper $35 00

FINKELSTEIN, DAVID H *Expression and the Inner* Cambridge Harvard, 2003 ix, 182 p Cloth $47 50

FINLAY, CHRISTOPHER J *Hume's Social Philosophy Human Nature and Commercial Sociability in A Treatise on Human Nature* New York Continuum, 2007 vii, 202 p $110 00 Cloth

FINN, DANIEL *The Moral Ecology of Markets Assessing Claims about Markets and Justice* New York Cambridge, 2006 182 p Cloth $65 00, paper $23 99

FINN, STEPHEN J *Thomas Hobbes and the Politics of Natural Philosophy* New York Continuum, 2006 ix, 192 p Cloth $120 00

FISCHER, JOHN MARTIN *My Way Essays on Moral Responsibility* New York, Oxford, 2006 260 p Cloth $45 00

FLANAGAN, OWEN *The Really Hard Problem Meaning in a Material World* Cambridge MIT, 2007 xiii, 288 p Cloth $27 95

FLEMING, BRUCE *The Aesthetic Sense of Life A Philosophy of the Everyday* Lanham, MD Rowman and Littlefield, 2007 vii, 180 p Paper $25 00, *The New Tractatus Summing Up Everything* Lanham, MD Rowman and Littlefield, 2007 ix, 128 p Paper $32 00, *Sexual Ethics Liberal vs Conservative* Lanham, MD Rowman and Littlefield, 2007 vi, 202 p Paper $20 00

FORBES, GRAEME *Attitude Problems An Essay on Linguistic Intensionality* New York Oxford, 2006 xi, 194 p Cloth $45 00

FORBES, HUGH DONALD *George Grant A Guide to His Thought* Toronto University Press, 2007 xii, 301 p Paper $27 95

FORREST, PETER *Developmental Theism From Pure Will to Unbounded Love* New York Oxford, 2007 199 p Cloth $55 00

FOX, WARWICK *A Theory of General Ethics Human Relationships, Nature, and the Built Environment* Cambridge MIT, 2007 x, 392 p Paper $28 00

FRANCES, BRYAN *Skepticism Comes Alive* New York Oxford, 2005 xii, 209 p Cloth $55 00

FRANKFURT, HARRY G *Demons, Dreamers, and Madmen The Defense of Reason in Descartes's Meditations* Princeton University Press, 2007 xviii, 264 p Paper $19 95

FRASER, IAN *Dialectics of the Self Transcending Charles Taylor* Charlottesville, VA Imprint Academic, 2007 viii, 205 p Paper $34 90

FRAZIER, BRAD *Rorty and Kierkegaard on Irony and Moral Commitment Philosophical and Theological Connections* New York Palgrave Macmillan, 2006 x, 249 p Cloth $69 95

FREEMAN, SAMUEL *Justice and the Social Contract Essays on Rawlsian Political Philosophy* New York Oxford, 2007 xii, 340 p Cloth $55 00

FRENCH, STEVEN *Science Key Concepts in Philosophy* New York Continuum, 2007 vii, 164 p Cloth $19 95

FRENCH, STEVEN, and DECIO KRAUSE *Identity in Physics A Historical, Philosophical, and Formal Analysis* New York Oxford, 2006 xv, 422 p Cloth $90 00

FREYDBERG, BERNARD *Imagination in Kant's Critique of Practical Reason* Bloomington Indiana UP, 2005 xii, 152 p Cloth $50 00, paper $19 95

FRICKER, MIRANDA *Epistemic Injustice Power and the Ethics of Knowing* New York Oxford, 2007 x, 188 p Cloth $49 50

FUDGE, ERICA *Brutal Reasoning Animals, Rationality, and Humanity in Early Modern England* Ithaca, NY Cornell, 2006 x, 224 p Cloth $45 00

FULLER, ROBERT C *Wonder From Emotion to Spirituality* Chapel Hill North Carolina UP, 2006 x, 188 p Cloth $24 95

FUMERTON, RICHARD *Epistemology* Malden, MA Blackwell, 2006 x, 145 p Cloth $49 95, paper $19 95

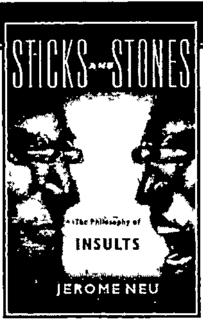
GAILUS, ANDREAS *Passions of the Sign Revolution and Language in Kant, Goethe, and Kleist* Baltimore Johns Hopkins, 2006 xvi, 222 p Cloth $60 00

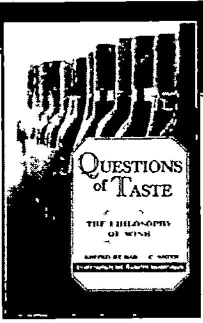
GANERI, JONARDON *Artha Meaning* New York Oxford, 2006 x, 258 p Cloth $35 00, *The Concealed Art of the Soul Theories of Self and Practices of Truth in Indian Ethics and Epistemology* New York Oxford, 2007 275 p Cloth $70 00

GARCIA, CARLOS E *Popper's Theory of Science An Apologia* New York Continuum, 2006 ix, 166 p Cloth $110 00

GARNSEY, PETER *Thinking about Property From Antiquity to the Age of Revolution* New York Cambridge, 2007 x, 267 p Cloth $90 00, paper $29 99

GARSON, JAMES W *Modal Logic for Philosophers* New York Cambridge, 2006 xv, 455 p Cloth $80 00, paper $34 99





New From OXFORD

Morality and Self-Interest
PAUL BLOOMFIELD
Ever since "Know Thyself" was inscribed at Delphi, Western philosophers have struggled to understand the relations between morality and self-interest. This edited volume of essays pushes forward one of the oldest and most important debates in philosophy.
2007 336 pp. cloth $99.00 paper $24.95

The Innate Mind
Foundations and the Future
Volume 3
Edited by PETER CARRUTHERS, STEPHEN LAURENCE, and STEPHEN STICH
This is the third volume of a three-volume set on The Innate Mind. By bringing together the top nativist scholars in philosophy, psychology, and allied disciplines, these volumes provide a compressive assessment of nativist thought and a definitive reference point for future nativist inquiry.
(Evolution and Cognition Series)
2008 456 pp., 3 halftones and 29 b/w line illus., 1 color line illus.
cloth $99.00 paper $29.95

Overcriminalization
The Limits of the Criminal Law
DOUGLAS HUSAK
Husak's goal in this book is to formulate a normative theory of criminalization that will allow us to distinguish which criminal laws are justified, and which are not – something he sees as essential in order to reverse the trend towards too many criminal laws.
2008 248 pp. $49.95

The Images of Time
An Essay on Temporal Representation
ROBIN LE POIDEVIN
The Images of Time is a philosophical investigation of the nature of time and the mind's ways of representing it. Robin Le Poidevin examines the ways in which we perceive time and change, the means by which memory links us with the past, the attempt to represent change and movement in art, and the nature of fictional time.
2007 192 pp. $55.00

Sticks and Stones
The Philosophy of Insults
JEROME NEU
In *Sticks and Stones*, philosopher Jerome Neu probes the nature, purpose, and effects of insults, exploring how and why they humiliate, embarrass, infuriate, and wound us so deeply. Neu draws upon a wealth of examples and anecdotes – as well as a range of views from Aristotle to Oscar Wilde to Katherine Hepburn and many others – to provide surprising answers to these questions.
2007 304 pp. $29.95

Questions of Taste
The Philosophy of Wine
BARRY SMITH
"The serious wine drinker seeking deeper knowledge and a more meticulous appreciation of his favorite beverage will take from these essays a glimpse into deep and subtle structures of the human mind."—Booklist
2007 288 pp. $27.95

OXFORD
UNIVERSITY PRESS
Prices are subject to change and apply only in the US. To order, please call 1-800-451-7556. In Canada, call 1-800-387-8020. Visit our website at www.oup.com/us

Young Epistemologist Prize

The **Rutgers Epistemology Conference** announces the sixth bi-annual prize for the best essay submitted to the Conference by a person with a Ph D obtained by the time of submission but not earlier than ten (10) years prior to the date of the conference Thus, for the Conference on May 8th/9th, 2009, the Ph D must not have been awarded prior to May, 1999 The prize winning essay will be presented at the bi-annual conference and will be published (along with the invited papers) in Philosophy and Phenomenological Research The winner of the prize will receive an award of $1,000 plus all travel and lodging expenses connected with the Conference

Details

The essay may be in any area of epistemology The essay ***must*** be limited to 5,000 words (not including footnotes and bibliography) and submitted on a PC-formatted disk along with three (3) printed copies The essays will be judged blindly and, thus, the three printed copies (but not the disk copy) must be prepared in such a way that the author's identity is completely masked ***Only*** papers submitted in that fashion will be considered Papers ***must*** be postmarked by November 15, 2008 The winner of the prize will be announced by February 15, 2009 Please include your **email address** in the cover letter accompanying your paper By submitting the essay, the author agrees to have it posted on the Rutgers Epistemology Conference Webpage and published in Philosophy and Phenomenological Research

Please send all papers (along with the disk and cover letter) to

The Young Epistemologist Prize
Philosophy Department
Rutgers University
26 Nichol Ave
New Brunswick, NJ 08901-2882

For more information about the Rutgers Epistemology Conference and the Young Epistemologist Prize, please go to the Conference Website

http://epistemology.rutgers.edu

THE JOURNAL OF PHILOSOPHY

Special Issues Available

Parts and Wholes — December 2006
On Bernard Williams — June 2003
Causation — April 2000
Jurgen Habermas and John Rawls — March 1995

The Dewey Lectures
Michael Dummett — *Truth and the Past*, 2002
Hilary Putnam — *Sense, Nonsense, and the Senses*
An Inquiry into the Powers of the Human Mind, 1994
Donald Davidson — *The Structure and Nature of Truth*, 1990
Amartya Sen — *Well-being, Agency, and Freedom*, 1985
John Rawls — *Kantian Constructivism in Moral Theory*, 1980
Ernest Nagel — *Teleology Revisited*, 1977
W. V. Quine — *Ontological Relativity*, 1968

The Woodbridge Lectures
John McDowell - *Having the World in View*
Sellars, Kant, and Intentionality, 1997

Subscriptions
Individuals $45.00
Students $20.00
Libraries and Institutions $100.00
Print and Online for Libraries and Institutions $180.00
Postage outside the United States $15.00

Special Offer
One complimentary issue of any of the above
issues with each new subscription
All issues are $5.00 plus $3.00 for shipping and handling
Send check or money order to
The Journal of Philosophy
Columbia University
1150 Amsterdam Avenue, Mail Code 4972
New York, NY 10027

www.journalofphilosophy.org

REVUE
INTERNATIONALE DE
PHILOSOPHIE

Direction: Michel Meyer

50, avenue Franklin Roosevelt, c.p. 188 / B-1050 Bruxelles (Belgique)
Tél. 32 2 650 25 61 / Fax 32 2 650 45 61
e-mail: tsalis@ulb.ac.be

3-2007 / volume 61 / n° 241

Deleuze

Ronald BOGUE
Editor's Introduction

Keith Ansell PEARSON
Overcoming the Weight of Man: Nietzsche, Deleuze, and Possibilities of Life

Paola MARRATI
Le nouveau en train de se faire: Sur le bergsonisme de Deleuze

Ronald BOGUE
The Art of the Possible

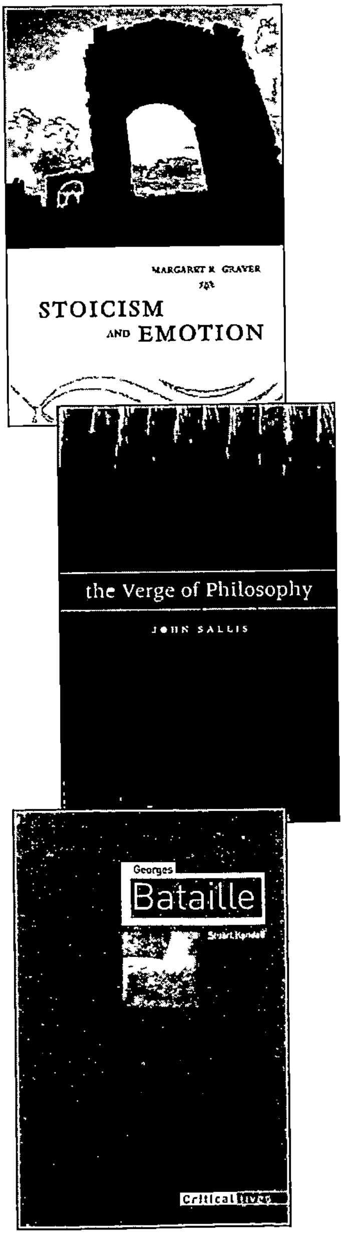
Elizabeth GROSZ
Deleuze, Bergson and the Concept of Life

Arnaud VILLANI
Gilles Deleuze ou la possibilité de vivre

Isabelle STENGERS
Penser la vie: le problème a changé

Prochains fascicules / Forthcoming

Levi-Strauss
Aesthetics
Philosophie de l'Esprit
La Barbarie
La Problèmatologie
Merleau Ponty



PHILOSOPHY

from Chicago

STOICISM AND EMOTION
Margaret R. Graver

"Margaret Graver, who has already done much to advance our understanding of the Stoic theory of emotions, here constructs a philosophically rich evaluation of the theory and its ethical implications. This elegantly argued volume deserves to become a milestone in the modern reengagement with Stoic thought."

David Sedley, Christ's College,
University of Cambridge

CLOTH $37.50

THE VERGE OF PHILOSOPHY
John Sallis

"In this brief yet highly engaging book, Sallis examines the limits and purpose of philosophy though the writings of Plato, Martin Heidegger, and longtime friend and interlocutor Jacques Derrida. ... Sallis has written a unique work that combines philosophical analysis with a heartfelt reflection on his friendship with Derrida."

Library Journal

CLOTH $25.00

New from Reaktion Books

GEORGES BATAILLE
Stuart Kendall

Georges Bataille was arguably the greatest influence on the poststructuralist revolution in twentieth-century thought and literature, yet few truly understand his work and legacy. Stuart Kendall translates the work and life of this renowned French writer, anthropologist, and philosopher into a concise yet informative biography.

PAPER $16.95

THE UNIVERSITY OF CHICAGO PRESS • www.press.uchicago.edu

New from Stanford University Press

The Discourse of the Syncope
Logodaedalus

JEAN-LUC NANCY

Nancy's classic study of the role of language in Kant demonstrates why the question of how to write philosophy, of *philosophical style*, is not just ancillary to critical philosophy but goes to the heart of the project of establishing human reason in its autonomy and freedom

Meridian Crossing Aesthetics
$21 95 paper $55 00 cloth

Psyche
Inventions of the Other, Volume II

JACQUES DERRIDA, Edited by PEGGY KAMUF and ELIZABETH G ROTTENBERG

"This monumental collection of essays shows Derrida at his brilliant best "
—Geoffrey Bennington, Emory University

Meridian Crossing Aesthetics
Mar 2008 $24 95 paper $65 00 cloth

Insister of Jacques Derrida

HELENE CIXOUS, Translated by PEGGY KAMUF

"*Insister of Jacques Derrida* is an indispensable, daring, heartfelt and moving book It presents a flawless, committed reading that is in the spirit of Derrida in its serious playfulness, its poetic sinuousness, its elegant reasoning and rhetoric while also being wholly in Cixous' own singular voice "
—Julian Wolfreys, Loughborough University

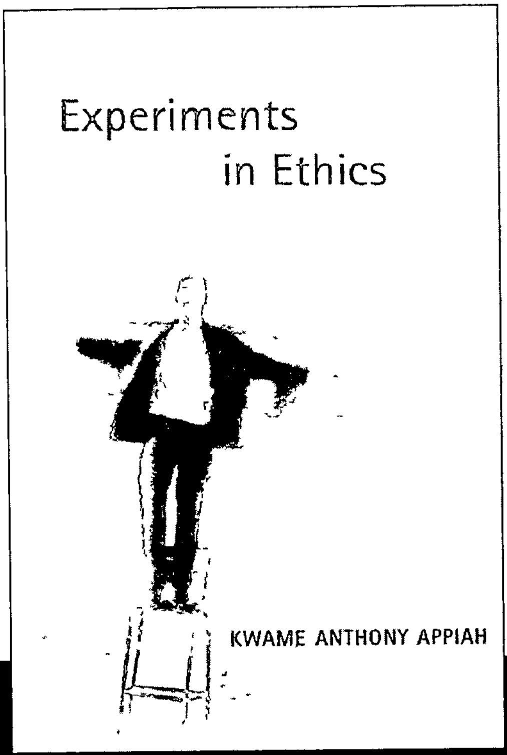
Edinburgh Univ Press
$21 95 paper $60 00 cloth

Stanford
University Press
800 621 2736
www sup org

You can now subscribe to the *Journal of Philosophy* using

PayPal

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org.**



"This dazzlingly written book argues for reconnecting moral philosophy with the sciences, both natural and social—and demonstrates that the reconnection, while in a sense overdue, reconnects philosophy with its ancient interest in empirical issues. Appiah's important argument promises to transform more than one field. It is not only wise and subtle; it is also inspiring."
—CASS SUNSTEIN, UNIVERSITY OF CHICAGO

"*Experiments in Ethics* is wonderful: concise but not breezy, clear but not simplistic, wide-ranging but focused, filled with wit and learning. It is an accessible, lively, and balanced introduction to empirical moral psychology that I recommend happily to philosophers and non-philosophers."
—WALTER SINNOTT-ARMSTRONG, DARTMOUTH COLLEGE
new in cloth

WWW.HUP.HARVARD.EDU **HARVARD UNIVERSITY PRESS**

The Search for Meaning

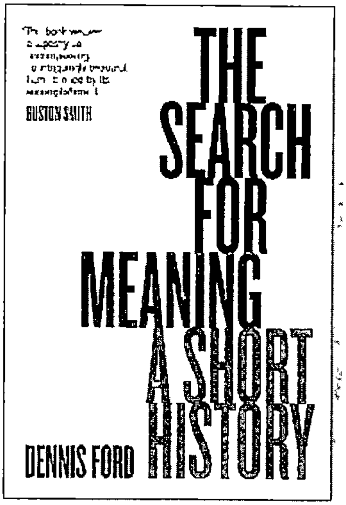
A Short History

DENNIS FORD

"Ford deserves to be congratulated for addressing a terribly important message so eloquently."

—**Jeffrey J. Kripal, author of *Esalen***

$24.95 hardcover



LUDWIG WITTGENSTEIN

Remarks on Colour

Edited by G. E. M. Anscombe
Translated by Linda L. McAlister
and Margarete Schattle

30TH ANNIVERSARY EDITION

This book comprises material on colour which was written by Wittgenstein in the last eighteen months of his life.

$16.95 paperback

Wittgenstein

Lectures and Conversations on Aesthetics, Psychology and Religious Belief

Edited by Cyril Barrett

40TH ANNIVERSARY EDITION

These notes should be of considerable interest to students of contemporary philosophy.

$14.95 paperback

Zettel

Edited by G. E. M. Anscombe
and G. H. von Wright
Translated by G. E. M. Anscombe

40TH ANNIVERSARY EDITION

"One of the most important volumes in the Wittgenstein literature." —***Choice***

$19.95 paperback

Presocratics

Natural Philosophers before Socrates

JAMES WARREN

"An excellent book and a pleasure to read." —**Steven Makin**

Ancient Philosophies

$50.00 hardcover, $19.95 paperback

Creationism and Its Critics in Antiquity

DAVID SEDLEY

"Sedley's treatment of ancient views on intelligent design will transform our current thinking."

—**Thomas Johansen, author of *Plato's Natural Philosophy***

Sather Classical Lectures

$29.95 hardcover

At bookstores or order (800) 822-6657 • www.ucpress.edu

UNIVERSITY OF CALIFORNIA PRESS

UNIVER
OF
CALCU

THE JOURNAL OF PHILOSOPHY

VOLUME CV, NUMBER 2
FEBRUARY 2008

page

55 *The Situation-Dependency of Perception* Susanna Schellenberg
85 *Is Realism Really Bad for You? A Realistic Response* Neera K. Badhwar
108 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Michael J Mooney, *Treasurer,* Liaquat Ahamed, *Other Trustees* Lee Bollinger, Leigh S Cauman, Barbara Gimbel, Kent Greenawalt, Lynn Nesbit, Daniel Shapiro

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at wwwjournalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2008

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2008 by the Journal of Philosophy, Inc ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CV, NO 2, FEBRUARY 2008

CENTRAL LIBRARY
UNIVERSITY OF CALCUTTA

THE SITUATION-DEPENDENCY OF PERCEPTION*

How can perception give us knowledge of the world? One of the challenges in answering this question is that we always perceive objects from a particular location, under particular lighting conditions, and similar features of the environment that determine the way objects are presented. Despite this fact, we can perceive the shape and color of objects.[1] We can perceive the shape of objects although only the facing surfaces are visible, and we can perceive two same-sized objects located at different distances to be the same size. Similarly, we can perceive the uniform color of a yellow wall although parts of it are illuminated more brightly than others, and we can recognize the sound of a cello regardless of whether it is played on a street or in a concert hall. More generally, subjects can perceive the properties objects have regardless of the situational features, although the way an object is presented always depends on the situational features.

When I speak of objects without further qualification, I mean mind-independent, external objects, such as cats and tables. Let us call the properties that an object has regardless of the situational features the *intrinsic properties* of the object. An intrinsic property of an object is a

---

* I am indebted to Keith Allen, Robert Audi, Stephen Engstrom, Alan Hájek, Benj Hellie, Anjana Jacob, James John, Uriah Kriegel, William Lycan, Mohan Matthen, Anders Nes, Gurpreet Rattan, Sydney Shoemaker, Declan Smithies, Evan Thompson, Timothy Williamson, Wayne Wu, and the members of the Australian National University philosophy of mind reading group for comments on a draft of this paper. Special thanks are due to David Chalmers, Anil Gupta, John McDowell, and Jonathan Schaffer. I am also grateful to members of audiences at the ANU, University of Oslo, University of Oxford, and Rutgers University—in particular Herman Cappelen, Bill Fish, Jason Stanley, Fiona Macpherson, and Kim Sterelny. Thanks finally to the Norwegian Research Council for generous funding while I was writing this paper.

[1] Of course, there are breakdown conditions. One cannot, for instance, tell whether a shape is a sphere or a cube, if one perceives it from sufficiently close up.

0022-362X/08/0502/55–84

property that does not depend on the object's relations to other individuals distinct from itself. The intrinsic properties that are of relevance to the current discussion are the perceivable intrinsic properties.[2] Let us call the features of the environment that determine the way an object is presented the *situational features*. The lighting conditions and the subject's location in relation to perceived objects are the situational features that are typically the most salient for the ways objects are presented—at least for human visual perception.

If the way an object is presented always depends on the situational features, then an explanation is needed for how one can have knowledge of the intrinsic properties of objects through perception. The aim of this paper is to contribute to such an explanation. The main thesis is that perceptual experience is necessarily situation-dependent. Call this the *situation-dependency thesis*. In short the idea is that the way an object is must not just be distinguished from the way it appears and the way it is represented, but also from the way it is presented given the situational features. I argue that the way an object is presented is best understood in terms of external, mind-independent, but situation-dependent properties that the object has given its intrinsic properties and the situational features [3] I will spell out the situation-dependency thesis in terms of the idea that perception represents situation-dependent properties To a first approximation, situation-dependent properties are extrinsic, mind-independent properties that are exclusively sensitive to and ontologically dependent on intrinsic properties and the situational features They play a pivotal epistemic role insofar as the perception of intrinsic properties is epistemically dependent on the representation of situation-dependent properties By arguing for this epistemic dependency, I aim to establish that the situation-dependency of perception is a desideratum for any account of perception

---

[2] For a discussion of intrinsic properties, see Brian Weatherson, "Intrinsic vs. Extrinsic Properties," in Edward Zalta, ed., *The Stanford Encyclopedia of Philosophy* (2006). In the context of the present discussion, I take for granted that colors are external, mind-independent, and intrinsic properties of surfaces, solids, or light. For a defense of colors as intrinsic properties, see Alex Byrne and David Hilbert, "Color Primitivism," in Ralph Schumacher, ed., *Perception and the Status of Secondary Qualities* (Boston: Kluwer, forthcoming). My argument is agnostic on the different possible ways of treating colors as mind-independent properties. So it is agnostic on views according to which colors are reflectance properties and so-called primitivist and naïve realist accounts of colors. For a defense of an understanding of colors as reflectance properties, see Mohan Matthen, "Biological Functions and Perceptual Content," this JOURNAL, LXXXV, 1 (January 1988): 5–27, and Byrne and Hilbert, "Color Realism and Color Science," *Behavioral and Brain Sciences*, XXVI, 1 (February 2003): 3–21. For critical discussions, see C.L. Hardin, *Color for Philosophers: Unweaving the Rainbow* (Indianapolis: Hackett, 1988), and Brian McLaughlin, "Color, Consciousness, and Color Consciousness," in Aleksander Jokiç and Quentin Smith, eds., *New Essays on Consciousness* (New York: Oxford, 2003), pp. 97–154. For a defense of a naïve realist view of color, see John Campbell, "A Simple View of Colour," in John Haldane and Crispin Wright, eds., *Reality, Representation, and Projection* (New York: Oxford, 1993), pp. 257–68. Colors understood in terms of dispositional properties are often considered to be mind-independent, for example, Sydney Shoemaker, "On the Way Things Appear," in Tamar Gendler and John Hawthorne, eds., *Perceptual Experience* (New York: Oxford, 2006), pp. 461–80. Similarly, views that analyze what it is for something to *be colored* in terms of what it is for something to *look colored* are sometimes considered to be compatible with the idea that colors are mind-independent, for example, Alva Noë, *Action in Perception* (Cambridge: MIT, 2004). Insofar as such accounts treat colors in terms of experiences of color, they are just a version of the view that colors are mind-dependent properties. For an excellent defense of this thesis, see Keith Allen, "The Mind-Independence of Colour," *European Journal of Philosophy* (forthcoming).

My argument has three parts First, I argue for the situation-dependency thesis Second, I show that perception of the intrinsic properties of objects is epistemically dependent on representations of situation-dependent properties Third, I argue that recognizing the situation-dependency thesis makes it possible to acknowledge that objects are presented a certain way, while holding on to the intuition that subjects directly perceive physical objects Before I embark on this project it will be helpful to motivate the situation-dependency thesis in light of standard views of perception

I BACKGROUND

Traditionally, the way an object is presented given the situational features has been downplayed or conflated with either the way the object appears to one (which is a matter of the phenomenology of perception) or with the way one represents the object (which is a matter of perceptual content) [4] Views that downplay the fact that objects are presented a certain way given the situational features insist that subjects perceive the intrinsic properties of objects directly If we perceive intrinsic properties directly, then it would seem that there is no need to appeal to what I call situation-dependent properties

---

[3] This use of the phrase that objects are presented in a certain way must be distinguished from the use of the phrase that picks out a subjective, sensational property of experience John McDowell writes of the determinateness with which shades of colors "are presented in one's visual experience"—*Mind and World* (Cambridge Harvard, 1994), p 56 What McDowell has in mind here is a particular phenomenon of the sensible qualities of experience McDowell's idea of something being presented in one's visual experience must be distinguished from the idea I am concerned with, namely the idea of the way objects are presented understood in terms of the mind-independent, situation-dependent properties of objects

[4] Appearances have in turn been understood in terms of the way an object is perceived, the way it is given in experience, the way it is presented in one's visual experience, or some other mind-dependent property of experience These notions differ substantially, but all concern the phenomenology of experience

Against such naive direct realist views,[5] I argue that if we take seriously *how* subjects gain knowledge through perception, we must acknowledge that the way an object is presented depends on the situational features I will argue that it follows from this that perceptual knowledge of intrinsic properties is epistemically dependent on representing the relevant situation-dependent properties If this is right, then the question of how subjects can have perceptual knowledge of objects cannot be answered by insisting that we perceive intrinsic properties directly

A second influential line of thought in the philosophy of perception is to understand the ways objects are presented given the situational features as mind-dependent This line of thought goes back to Locke, Berkeley, and Hume Hume writes "[T]he table, which we see, seems to diminish, as we remove farther from it but the real table, which exists independent of us, suffers no alteration it was, therefore, nothing but its image, which was present to the mind "[6] By presupposing that the appearance of an object is a case of an illusion, Hume is led to conclude that appearances are mind-dependent items of sorts The thought that there is something present to the mind that is distinct from the perceived object has motivated different versions of phenomenalism, sense-data theories, and indirect realism [7] The details of these views differ radically, in particular with respect to the ontological and epistemological status of that which is directly present to the mind What the views in this tradition have in common is the thesis that what is directly present to consciousness is a mind-dependent property or object rather than a property of the physical object that the perception is of [8]

---

[5] It is important to distinguish "naïve direct realism" from "naïve realism" as defended, for example, by Michael Martin, "The Reality of Appearances," in Mark Sainsbury, ed , *Thought and Ontology* (Milan, Italy Franco Angeli, 1997), pp 81–106 The former is a thesis about the direct object of perception, the latter is a thesis about the mind-independence of the objects, properties, and relations that we perceive

[6] Hume, "An Enquiry concerning Human Understanding," in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, L A Selby-Bigge and P H Nidditch, eds (New York Oxford, 1975), p 152

[7] H H Price famously argues that sense-data are mind-dependent, nonphysical objects See his *Perception* (London, UK Methuen, 1950, first published 1932) Frank Jackson defends the existence of nonphysical sense-data on the basis of linguistic analysis See his *Perception* (New York Cambridge, 1977) He interprets the claim "object *x* looks red" as having the underlying form, "Subject *S* sees a red sense-datum belonging to *x*" More recently, he has distanced himself from his unabashed sense-data view See his *From Metaphysics to Ethics* (New York Oxford, 1998) I follow A D Smith in treating sense-data theories as a version of indirect realism See his *The Problem of Perception* (Cambridge Harvard, 2002)

[8] This thesis is compatible with the claim that the intentional object of perception is an external, mind-independent object and with the claim that perceptual content represents (possibly among other things) external and mind-independent objects

Against the views in this tradition, I argue that at least part of what is typically understood as mind-dependent should be understood in terms of the situation-dependency of perception In short, the idea is that the ways objects are presented are best understood as external, mind-independent, but situation-dependent properties of the object Understanding the ways objects are presented as external and mind-dependent makes it possible to distinguish between the subjective aspect of perception, on the one hand, and its objective, but situation-dependent aspect on the other

The distinction between the subjective and objective properties of perception has been understood in many different ways, but all concern the dependence or independence of perception on the subject's sensibility or cognition In the idealist tradition, the distinction is understood to mark the extent to which subjects perceive reality independently of their sensibility or cognition Insofar as perception is subjective, perception is understood to be constitutive of its object In the phenomenological tradition, the distinction is understood to mark the extent to which the object of phenomenal consciousness is mind-dependent

Phenomenalism, indirect realism, and direct realism differ with respect to what constitutes the *object* of perception The question of what perception is of and what perception represents must be distinguished from the question of how the relationship between the representational content and the phenomenology of experience is conceived The argument presented here is agnostic on the answer to the second question If the argument holds at all, it will hold regardless of how the debate on the relation between the representational content and the phenomenology of experience is resolved The present discussion does however relate to that debate insofar as I argue that at least part of what is often conceived of *only* in terms of the phenomenology of experience should be thought of, at least in part, in terms of representations of external, mind-independent properties of objects This by no means amounts to a denial of nonrepresentational appearance properties As I will argue in the next section, the thesis that perception involves representations of situation-dependent properties is compatible with these properties being accompanied by appearance properties

II THE ARGUMENT FOR THE SITUATION-DEPENDENCY THESIS

Phenomenalists and indirect realists treat the ways objects appear as something mind-dependent There are no doubt mind-dependent ways that objects appear However one understands these mind-dependent ways that objects appear, I will argue that experience represents the *mind-independent* ways that objects are presented given the situational features The way an object is presented can be understood strictly externally, namely as determined by the object's intrinsic properties and the situational features [9] For the visual perception of size and shape, the perceiver's location in relation to the perceived object is the crucial situational feature that determines how the object is presented For the perception of color, the lighting conditions and color context are among the crucial situational features

If the way an object is presented is understood as external and mind-independent, it can be understood in terms of the properties the object has given the situational features, that is, the object's situation-dependent properties [10] *Situation-dependent properties* are (nonconstant) functions of the intrinsic properties of the object and the situational features This means that fixing the intrinsic properties and the situational features fixes the situation-dependent properties Furthermore, situation-dependent properties are ontologically dependent on and exclusively sensitive to intrinsic properties and situational features [11]

It will be helpful to consider some examples Take the cup on my table It is presented in a certain way given my location One side is closer than the other, one part faces away from me Its shape is presented in an egocentric frame of reference, which in turn means that the object and its parts are presented as standing in specific spatial relations to me The way the cup is presented to a location is on the suggested view an external and mind-independent, albeit situation-dependent property of the world Any perceiver occupying the same location would, *ceteris paribus,* be presented with the cup in the very same way Similarly, the way the color of the cup is presented is the way its intrinsic color property manifests itself given the current lighting conditions If the intrinsic color property is conceived of as a reflectance property, then the correlating situation-dependent property will be a wavelength emittence property

---

[9] This thesis draws on several well-known concepts Noë's (*op cit*) notion of P-properties and Christopher Peacocke's notion of a scene—*A Study of Concepts* (Cambridge MIT, 1992), see especially chapter 3 As I show below, situation-dependent properties can be understood as the properties constituting scenes I discuss Noë's view in my "Action and Self-Location in Perception," *Mind*, cxvi, 463 (July 2007) 603–32 In brief, Noë understands P-properties as projections on a plane perpendicular to the subject's line of sight Although he insists that they are objective, mind-independent properties, he analyzes them in terms of how things look to perceivers, more specifically in terms of what he calls phenomenal objectivism and in terms of the psychological notion of a visual field In this respect, they are the very same kind of properties that James J Gibson refers to as visual field properties and Peacocke refers to as sensational properties See Gibson, *The Perception of the Visual World* (Boston Houghton Mifflin, 1950), and Peacocke, "Sensational Properties Theses to Accept and Theses to Reject," in *Revue Internationale de Philosophie*, in a special issue on the Philosophy of Mind edited by Joëlle Proust (forthcoming) For a defense of this thesis, see Allen, "Being Coloured and Looking Coloured" (manuscript) Matthen argues for an idea that is similar to Noë's—*Seeing, Doing, and Knowing* (New York Oxford, 2005) According to Matthen, a surface is presented as it is—color-wise—in part because of a property it has, its color, and in part because of the condition it is in In contrast to P-properties, situation-dependent properties are not analyzed in terms of how things look Furthermore, they are not properties of the visual field and are not two-dimensional geometrical projections However, as I will argue, recognizing situation-dependent properties is compatible with there being properties that capture how things look

[10] An alternative label might be "relational properties" But relational properties have been understood as properties that involve relations to a perceiver's mind For discussion, see Evan Thompson, *Colour Vision* (New York Routledge, 1995) Since the properties I am introducing are a function only of external, mind-independent features it is important to distinguish them clearly from such relational properties

[11] Thanks to Alan Hájek for helpful discussions about different ways of understanding relations of ontological dependence

It is conceivable that two objects with different intrinsic properties can have the very same situation-dependent property given suitable situational features So it is conceivable that, say, a blue couch under yellow lighting conditions has the very same situation-dependent color property as a green couch under standard lighting conditions Furthermore, the thesis that there is a distinction between the situation-dependent properties and the intrinsic properties of an object is compatible with the existence of cases in which the situation-dependent and intrinsic properties of an object match Say I am looking at the round rim of a cup from directly above From such a location, the rim of the cup is presented as round But even in this case, there is a metaphysical distinction between the two properties The important point is that given the spatiotemporal nature of perception,[12] there is a distinction between the way an object is presented and the way it is regardless of the situational features In other words, there is a distinction between the situation-dependent and the intrinsic properties of an object

The idea of situation-dependent properties can be understood in terms of Christopher Peacocke's notion of a scene A scene is "the volume of the real world around the perceiver at the time of the experience, with an origin and axes in the real world fixed in accordance with the labeling of the scenario "[13] The scenario is a way of locating objects, properties, and relations in relation to a labeled origin and axes The elements of a scene can be understood as situation-dependent spatial properties An object is presented in different ways to different locations in a world and it is an objective mind-independent feature of the object that it is so presented Since situation-dependent spatial properties are location-relative properties rather than perceiver-relative properties, any perceiver occupying the same location is, *ceteris paribus*, presented with the very same situation-dependent spatial properties.

---

[12] I elaborate on this spatiotemporal nature of perception in section III

[13] Peacocke, *A Study of Concepts,* p 64

The *distinction* between a scene and a world that is not location-relative is, however, only helpful to bring out the contrast between intrinsic *spatial* properties and the correlating situation-dependent properties. It is not helpful to bring out the contrast between intrinsic colors or sounds and the correlating situation-dependent properties. The reason is that the situational features relevant for perceiving spatial properties is location-relative (for instance, the distance from the perceiver and the orientation relative to the perceiver), while the situational features relevant for perceiving color properties are not (for instance, the lighting conditions). Lighting conditions are present in the scene and the world that is not location-relative. Therefore, the distinction between intrinsic color properties and the correlating situation-dependent properties cannot be analyzed in terms of the distinction between scenes and the world that is not location-relative. What we need is a distinction between scenes and the intrinsic properties in the relevant part of the uncentered world. That is just the distinction between situation-dependent and intrinsic properties.[14]

The very same two aspects that are involved in veridical perception can be argued to be involved in cases of illusion and hallucination, at least if the illusion or hallucination is of a seemingly colored and three-dimensional space-occupier. When one hallucinates an elf, one has a seeming point of view in relation to the elf. One can distinguish between the way the elf seems to fill out space from the way it seems to be presented given one's location. So the distinction between situation-dependent and intrinsic (hallucinated) properties holds for cases of illusion and hallucination in the very same way as it holds for cases of accurate perception.

Thinking about the ways objects are presented in terms of situation-dependent properties makes it possible to acknowledge that as the situational features change, perceptual content remains the same with regard to the intrinsic properties of objects, but changes with regard to the situation-dependent properties. Furthermore, by understanding the ways objects are presented in terms of situation-dependent properties, the difference between how objects are presented and how one perceives them to be can be clearly brought out as a difference in *what* is perceived, and thereby distinguished from any differences in *the ways* in which an object, property, or relation is perceived The point is to distinguish

---

[14] With certain qualifications, situation-dependent properties could also be understood as what constitutes a centered world or in terms of centered properties. The *distinction* between a centered and an uncentered world is, however, only helpful to bring out the contrast between intrinsic *spatial* properties and the correlating situation-dependent properties. It is not helpful to bring out the contrast between intrinsic colors or sounds and the correlating situation-dependent properties.

(1) the way an object is presented given the situational features

from any

(2) mind-dependent objects or properties [15]

The distinction between situation-dependent and intrinsic properties is not a distinction between the properties objects appear to have and the properties those objects actually have Nor is it a distinction between the properties objects appear to have and the properties our experience represents objects as having Since both intrinsic and situation-dependent properties are external and mind-independent, the distinction between them does not run parallel to the distinction between the properties of experience and the properties of objects The distinction lies rather within the realm of external, mind-dependent properties of objects

I am not arguing that every phenomenological difference should be understood in terms of a representational difference Take Ernst Mach's example of perceiving a shape from the same angle once as a square and once as a diamond The difference in perception is due to a difference in the way it is perceived As Peacocke has argued, the way it is perceived contributes to what it is like to have the experience and is not captured in the representational content of perception [16] In such cases, there is no external difference to be represented and so the difference in phenomenology cannot be explained in terms of a difference in representations of mind-independent properties or objects

I am arguing only that what is often understood *solely* in terms of mind-dependent properties or objects should be understood at least in part in terms of situation-dependent properties When one changes one's location in relation to a shape, the way the shape is presented changes, that is, something external and mind-independent changes, namely the situation-dependent property of the object In contrast to Mach's case, the change in the experience due to the change in situational features can be explained with regard to the situation-dependency, rather than the subjectivity of perception. To a first approximation the subjectivity of experience can be understood as any aspect of experience that is a matter of mind-dependent properties or objects of the experience rather than the external, mind-independent objects and properties that the experience is of. The point of distinguishing the situation-dependency from the subjectivity of perception is to contrast the properties of external objects from the properties of experience and the mind-dependent objects of experience.

---

[15] The mind-dependent objects or properties can be understood for instance in terms of sensational properties, appearance properties, or any number of mind-dependent ways that an object appears to a subject

[16] See, for instance, his *A Study of Concepts*, pp 74–77, and also his "Does Perception Have a Nonconceptual Content?," this JOURNAL, XCVIII, 5 (May 2001) 239–64, p 253f

Insisting that the ways objects are presented are mind-independent is of course not to deny that representations of these ways are mind-dependent. But it is crucial to distinguish

(3) representations of the way an object is presented given the situational features

from how the object appears to be and from the way the object is presented. How one represents objects as being presented is dependent on one's sensory system. Representations have been understood in many different ways. The present argument holds for any understanding that is committed to the most basic principle of characterizing perceptual content in terms of representations, namely, that there is a distinction between the case in which things are as they are represented to be and the case in which things are otherwise. In other words, the only principle that the use of representations here is committed to is that representations have accuracy conditions.[17] This commitment is neutral on the question of whether perceptual content is conceptual or nonconceptual, whether perceptual content is propositional or image-like, or whether representational properties are pure or impure.[18]

In the interest of generality, I will not commit myself to a particular way of thinking about representations other than representations having accuracy conditions. It is crucial, however, to be specific about *what* is represented when a subject has a scene in view. What one represents cannot simply reflect how objects are presented. It cannot be right that subjects, say, represent everything that they have in view. A myopic perceiver will represent less detail than someone with excellent vision given the very same scene, and a cat will represent something different from a human because its sensory apparatus is set up differently. Thus, what a subject represents must depend on what is perceptually available to her. What is perceptually available to a subject is determined by what could be perceptually discerned by her, which in turn is determined by her perceptual system. The number of leaves on a tree is, for example, not perceptually available to most humans, since our perceptual system does not have the capacity to perceive numbers of similar objects far beyond five. There is a further complication. Subjects are not necessarily aware of everything that is perceptually available to them. Say I see a tree swaying in the breeze. I might not be aware of the specific movement of a particular leaf fluttering in the breeze or a bug on one of the twigs, although these are details that are perceptually available to me and thus details of which I could be aware. So what is perceptually available must be determined dispositionally: for something to be perceptually available is for it to be something of which a subject can be perceptually aware given her perceptual system.

---

[17] In the context of this paper, I take for granted that experience is subject to accuracy conditions. This idea has been denied by adverbialists, such as Roderick Chisholm—see his *Perceiving* (Ithaca: Cornell, 1957). More recently, the thesis has been argued against by Charles Travis, "Silence of the Senses," *Mind*, CXIII, 449 (January 2004): 57–94, and by Bill Brewer, "Perception and Content," *European Journal of Philosophy*, XIV, 2 (August 2006): 165–81.

[18] Pure representational properties are properties of representing a content, while impure representational properties are properties of representing a content in a certain way. For a defense of this distinction, see Chalmers, "The Representational Character of Experience," in Brian Leiter, ed., *The Future for Philosophy* (New York: Oxford, 2004), pp. 153–81.

The important point for the present argument is that the differences in our perceptual systems affect only what we *represent* of a perceived scene. It does not affect the situation-dependent properties. It is crucial to distinguish both between what is there to be picked up and what one's perceptual system can pick up, as well as between what one's perceptual system in fact picks up and what one is perceptually conscious of. *How* one represents situation-dependent properties and *what* one represents when one has a particular scene in view depends on one's perceptual system. By contrast, the situation-dependent properties are there to be picked up regardless of whether one is the kind of perceiver that can pick them up and regardless of whether one actually does pick them up. In this respect, situation-dependent properties are just like intrinsic properties.

It will be helpful to work through an example. Consider Peacocke's example of two same-sized trees located at different distances from a subject. Peacocke analyzes the difference between the actual and the apparent sizes of the trees in terms of "a duality of representational properties and properties of the two-dimensional visual field "[19] He identifies the properties of the two-dimensional visual field as sensational properties So according to Peacocke, the difference between the perceptions of each tree is a difference in sensations, not a difference in representations

By recognizing situation-dependent properties, at least part of what Peacocke understands as a difference in *sensations* can be reduced to a difference in *representations* of situation-dependent properties So Peacocke's duality can be thought of as a duality within the representational properties Such an approach allows one to think of the apparent difference in size not as an aspect of the visual field, but rather as an aspect of the world as it is presented to the perceiver's location I am not denying that there is a visual field and I am not denying that the representations of situation-dependent properties could be matched by sensational properties I am arguing that Peacocke's trees can be explained without appeal to the psychological notion of a visual field They can be explained with regard to the world as it presents itself to the perceiver's location, that is, with regard to external, mind-independent, but situation-dependent properties of objects Since the situation-dependent properties are actual properties of the object given the situational features, one can say that there is a difference between the perception of the two trees with regard to the properties represented, and thus a difference in the externally individuated content of perception In the veridical case, experience represents the trees as having the same intrinsic size properties, but as having different situation-dependent size properties Once representations of situation-dependent properties are recognized, the apparent difference in size between the trees can be understood as a difference in situation-dependent properties So once situation-dependent properties have been introduced, there is no need to appeal to sensational properties in the subject's visual field to explain the apparent difference in size of the two trees

Now one could object that this leads to the problem that perception will represent the trees as both the same size and as different in size

(a) (*Tree*$_1$ and *Tree*$_2$ are different in size, *Tree*$_1$ and *Tree*$_2$ are the same size)

This cannot be the right way of characterizing the content since it would render the content incoherent Since the situation-dependent properties will differ from the intrinsic properties in any standard case of perception, most perceptual contents (so characterized) will exhibit a similar incoherence

---

[19] Peacocke, *Sense and Content Experience, Thought, and Their Relations* (New York Oxford, 1983), p 13

Since, however, the sense in which the two trees are different in size is distinct from the sense in which they are the same size, the content in the case at hand is not incoherent When the two components of perceptual content are assessed, they are assessed relative to different aspects of the object (the object given the situational features, the object regardless of the situational features) The incoherence of (a) is just a way of describing the puzzle of perception that we are trying to come to grips with one can perceive an object's intrinsic properties although the way an object is presented depends on the situational features By showing that subjects represent both intrinsic and situation-dependent properties, it is possible to acknowledge that there is a sense in which perceiving the two trees as different in size is accurate they really are presented as different in size given the subject's location in relation to the trees

These considerations bring out a second problem with the way (a) characterizes the content in question The two trees do not look to be both the same and different in size They look to be the same size due to the fact that the closer tree is presented as larger than the tree that is further away So the situation-dependent properties and intrinsic properties are not on par One could try to account for the difference between the two properties by including more than the outcome of the function

(b) (*Tree*$_1$ and *Tree*$_2$ are different in size given the situational features, *Tree*$_1$ and *Tree*$_2$ are the same size)

This suggestion is similar to Gilbert Harman's suggestion Harman argues that the "feature of a tree from here is an objective feature of the tree in relation to here, a feature to which perceivers are sensitive and which their visual experience can somehow represent things as having from here "[20]

The problem with the content characterized by (b) is that it involves more information than is required The thesis that perception represents situation-dependent properties as well as intrinsic properties does not require that the *situational features* under which an object is perceived be represented The specific angle at which one perceives a table or the precise brightness of the light are features that are arguably not perceptually available, at least not to most human perceivers More importantly, subjects can perceive the trees to be different in size given their location, without being aware of their location in relation to the trees All that the situation-dependency thesis is committed to is that subjects represent *situation-dependent properties* and so the *effect* of situational features This idea involves no commitment to saying that subjects represent *situational features*

---

[20] Harman, "The Intrinsic Quality of Experience," in James Tomberlin, ed , *Philosophical Perspectives Action Theory and Philosophy of Mind*, Volume 4 (Northridge, CA Ridgeview, 1990), pp 31–52

Another way of bringing out the issue is with regard to the difference between relational and monadic properties Situation-dependent properties can appear to be monadic properties But a property can appear to be monadic and nonetheless reveal itself to be relational on reflection[21] Although situation-dependent properties are relational properties insofar as they are a function of intrinsic properties and situational features, they need not be represented as relational properties Moreover the fact that they are relational properties need not reveal itself in the phenomenology of perception One might want to take these considerations into account by characterizing the content in the following way

(c) (*Tree*$_1$ and *Tree*$_2$ are situation-dependently different in size, *Tree*$_1$ and *Tree*$_2$ are intrinsically the same size)

This proposal is similar to Michael Tye's[22] Tye argues that "experience represents the nearer tree as having a facing surface that differs in its viewpoint-relative size from the facing surface of the further tree, even though it also represents the two trees as having the same viewpoint-independent size" (*ibid*, p 124)

The problem with this characterization of the content is that the content represents *whether* a property is a situation-dependent property or an intrinsic property Although human perceivers might typically be able to distinguish, say, the intrinsic shape of an object from the correlated situation-dependent property, it is easy to imagine a situation in which a subject is not able to tell whether a perceived property is intrinsic or situation-dependent Tye goes on to argue that the nearer tree is represented *as* being larger *from here* This idea implies that perceivers are aware not only whether a property is a situation-dependent property or an intrinsic property, it implies that perceivers are aware of the situational features

---

[21] For a discussion of such features of properties, see Anil Gupta, "Meaning and Misconceptions," in Ray Jackendoff et al, eds, *Language, Logic, and Concepts* (Cambridge MIT, 1999), pp 15–41

[22] Tye, "Perceptual Experience is a Many-Layered Thing," in E Villanueva, ed, *Philosophical Issues*, Volume 7 (Atascadero, CA Ridgeview, 1995), pp 117–26

As I argued above this is cognitively too demanding a requirement for perception

These problems can be avoided by characterizing the content in the following way

(d) ($\textit{Tree}_1$ and $\textit{Tree}_2$ are the same size′, $\textit{Tree}_1$ and $\textit{Tree}_2$ are different in size″)

Metaphysically speaking the single primed property is a situation-dependent property and the double primed property is an intrinsic property But perception need not represent these metaphysical facts

An analogous analysis can be given for a case involving color properties Consider a subject who is perceiving a blue couch in bright light Due to the bright light, the couch is situation-dependently green The subject's perceptual content can be characterized in the following way

(e) (that couch is blue′, that couch is green″)

This way of characterizing the content makes it possible to acknowledge that the situation-dependency thesis involves no commitment to subjects knowing whether a property is a situation-dependent property or an intrinsic property

There are many possible ways of understanding the way situation-dependent properties are reflected in the phenomenology of perception My argument is agnostic on how this issue is conceived I aim to establish that whatever stance is taken toward the way *intrinsic* properties are reflected in phenomenology should be taken towards the way situation-dependent properties are reflected in phenomenology Whatever differences there are in phenomenology are arguably psychological differences For human perceivers, intrinsic properties are usually more salient than situation-dependent properties Therefore, they are typically in the foreground of our perceptual phenomenology [23]

---

[23] David Chalmers discusses many possible ways of thinking about the phenomenology of the perceptual experience of a wall that is partially in shadow See his "Perception and the Fall from Eden," in Gendler and Hawthorne, eds , *Perceptual Experience*, pp 49–125, in particular pp 84–89 His distinction between simple and complex phenomenology is particularly salient for the present discussion On the simple view, only what I call situation-dependent properties are reflected in the phenomenology of an experience On the complex view, the experience represents an intrinsic property in conjunction with a situational feature As I have argued, this involves more information than required The specific angle at which one perceives a table or the precise brightness of the light are features that are arguably not perceptually available, at least not to most human perceivers I take it to be plausible that typically both situation-dependent properties and intrinsic properties are reflected in

This analysis is congenial to William Lycan's argument[24] that the apparent difference in the sizes of Peacocke's trees is represented in experience. However, it differs from his analysis insofar as my thesis is not understood to imply that something other than mind-independent properties, objects, and relations are represented. Lycan argues that perceptual representations have at least two layers. One is concerned with external, *mind-independent* physical objects, properties, and relations, the other is concerned with *mind-dependent* intentional objects and relational properties, namely colored shapes. "We do visually represent the trees as being of the same size etc., but we do that by representing colored shapes and relations between them. Some of the shapes—in particular those corresponding to the trees—are represented as being larger shapes than others, as occluding others, and so forth. As with all intentional objects, it does not follow that there are any actual things that have such relational properties" (*ibid.*, p. 95). Lycan introduces a kind of mind-dependent intentional object to account for the two trees appearing different in size. He mirrors Hume's motivation for introducing such objects when he writes: "in Peacocke's example there are a veridical road-and-same-sized trees presentation and an illusory different-sized-tree-shapes presentation made in the same visual experience. The experience has one truth-condition that would be satisfied by the apparent different-sized shapes, and another that is satisfied by the same-sized tress. It happens that the former is actually not satisfied and in that sense there is an illusion" (*ibid.*, p. 129). On Lycan's view, a subject's experience of the trees represents two same-sized trees as well as two intentional objects, one larger than the other. He argues that the representation of such intentional objects, say, a colored shape, need not imply that the colored shape exists. The shape is an object, but it is not an object that actually exists (*ibid.*, p. 90).

Recognizing situation-dependent properties makes it possible to acknowledge that there is nothing illusory about representing the trees to be presented as different in size given the situation. Indeed given the situation, representing the trees to be presented as different in size is *correct*. Since situation-dependent properties are actual

---

the phenomenology of an experience. So my view suggests neither Chalmers's simple or complex view about phenomenology, but rather a third view. Call it the multiplex view. I will save a detailed discussion of the role that situation-dependent properties play in phenomenology for a future occasion.

[24] Lycan, "Layered Perceptual Representations," and "Replies to Tomberlin, Tye, Stalnaker, and Block," in Villanueva, ed. (*op. cit.*), pp. 81–100, 127–42.

properties of the environment, the thesis that experience represents them involves no commitment to saying that experience represents anything other than mind-independent properties, objects, and relations Recognizing situation-dependent properties is, however, compatible with Lycan's thesis that experience represents intentional objects Similarly, one can recognize situation-dependent properties and nonetheless be a phenomenalist or an indirect realist On such a view, there will be mind-dependent objects or properties that stand in representation relations to *both* intrinsic properties and to situation-dependent properties

But although formally consistent with phenomenalism and indirect realism, recognizing situation-dependent properties undermines at least one traditional motivation for such views As I argued earlier, one traditional motivation is that when a subject, say, moves away from a table something becomes smaller Since it is not the table that becomes smaller, what is directly present to the mind must be something other than the table If one recognizes situation-dependent properties, no appeal to mind-dependent objects or properties is necessary to explain how there can be a way that objects look that is not accounted for by representing their external, mind-independent properties Indeed, such cases are explained better with appeal to representations of two kinds of mind-independent properties when one moves away from the table the situational features change while the intrinsic properties remain the same The change in situation-dependent properties accounts for the change in the experience If such cases are better explained with appeal to two kinds of mind-independent properties, then once situation-dependent properties are introduced, the mind-dependent objects or properties are obsolete

There are, however, other reasons to introduce mind-dependent objects or properties One is that mind-dependent sensational properties are what make an experience an experience[25] A second is to account for the possibility that hallucinatory and veridical experiences are phenomenologically indistinguishable[26] A third is to account for the possibility of spectrum inversion Sydney Shoemaker introduces appearance properties to reconcile the possibility of an intentionalist account of perception with phenomenal spectrum inversion without misrepresentation.[27] Take two subjects, one of which is spectrum inverted. Both subjects correctly represent a red tomato as red, but the phenomenology of their experiences differs. While one subject experiences the phenomenology that I experience when I am acquainted with red objects, the other subject experiences the phenomenology that I experience when I am acquainted with green objects. Representations of appearance properties are introduced to account for the difference in phenomenology. The subjects represent the tomato as having different appearance properties, but neither of them is misrepresenting.[28] Appearance properties are either occurrent or dispositional properties. An occurrent appearance property is the property of currently producing a certain experience in a perceiver, a dispositional appearance property is a disposition to cause experiences of a certain sort in a kind of perceiver. According to Shoemaker, the phenomenal character of experience consists in the representation of either kind of appearance property.[29]

---

[25] For a defense, see Peacocke, "Sensational Properties "

[26] I am following Timothy Williamson in using the expression of phenomenological indistinguishability to pick out an epistemic notion—*Identity and Discrimination* (Cambridge Blackwell, 1990) On this understanding, for two experiences to be phenomenologically indistinguishable just is for the subject not to be able to distinguish them

There are two parameters on which situation-dependent properties differ from appearance properties. While an appearance property "is such that something one sees has it just in case it appears to one to have it,"[30] a situation-dependent property is a property that a perceived object can have without one being conscious of it having the property. Put more simply, situation-dependent properties are properties that objects *actually* have given the situational features regardless of what any potential perceiver could pick up given its sensory capacities. So while appearance properties are relative to the sensory capacities of a potential perceiver, no reference to a perceiver (or a potential perceiver) is necessary to specify the situation-dependent properties of an object.

The second difference concerns the question of how many kinds of properties are in the external, mind-independent realm. On Shoemaker's view, experience of colors requires representing properties other than colors. Subjects perceive colors by perceiving properties distinct from them. As a consequence, each color experience ascribes two different properties (*ibid.*, p. 467). Michael Thau defends a view in opposition to Shoemaker, according to which subjects perceive colors by ascribing only one property.[31] Thau calls it a nameless intrinsic property. But like Showmaker, Thau argues that perception represents a property that is distinct from the color of the object that tells the perceiver about the color of the object. In contrast to both Shoemaker and Thau, I am arguing that experience of intrinsic colors requires representing two kinds of mind-independent, external properties of the object: intrinsic properties and situation-dependent properties. While appearance properties are relative to what a particular kind of perceiver can pick up, both intrinsic properties and situation-dependent properties are what are there to be picked up regardless of what a particular perceiver can pick up.

---

[27] Shoemaker used to call these properties "phenomenal properties" but has changed his terminology to "appearance properties." See his "Introspection and Phenomenal Character," *Philosophical Topics*, xxviii (2001): 247–73.

[28] For a critical discussion of Shoemaker's appearance properties as candidates for reconciling the possibility of spectrum inversion and intentionalism, see Andy Egan, "Appearance Properties?," *Noûs*, xl, 3 (September 2006): 495–521.

[29] Recently, Shoemaker has argued that the phenomenal character sometimes does not consist in the representation of appearance properties, but rather in the representation of one or another of the aspects of perceivable properties that he calls qualitative aspects. See his "On the Way Things Appear." But like appearance properties, qualitative aspects are characterized in terms of how they affect or are disposed to affect the phenomenal character of experiences.

[30] Shoemaker, "On the Way Things Appear," p. 465.

So what I am concerned about is orthogonal to Shoemaker's concerns. Indeed one can embrace situation-dependent properties along with Shoemaker's appearance properties or any other kind of mind-dependent properties. The mind-dependent properties will stand in representation relations both to intrinsic properties and to situation-dependent properties.

But what is the advantage of recognizing situation-dependent properties if one already has mind-dependent properties in place? There are several advantages, which I will discuss in turn. One advantage is that it makes it possible to recognize that more of experience is subject to accuracy conditions than is traditionally acknowledged. For an experience to be accurate it must not just get the intrinsic properties of an object right, but also its situation-dependent properties. The *ways* objects are presented given the situational features makes a difference to which perception it is. Perceptions are individuated more finely than by objects and their intrinsic properties; they are individuated also by situation-dependent properties.[32] If perceptions are not just individuated by the objects they are of, but by the way those objects are presented, then individuation of perceptual content necessarily requires appeal to facts about the situation in which the object is presented. If this is right, then perception involves representations not just of an object's intrinsic properties, but also of its situation-dependent properties. As a consequence, to assess whether an experience is accurate is not only to assess whether the right intrinsic properties are represented, but also to assess whether the right situation-dependent properties are represented. This is just to say that the satisfaction of the accuracy conditions depends on accurately representing both the object's intrinsic properties and its situation-dependent properties. If situation-dependent properties are external and mind-independent properties, then they can be represented and misrepresented in just the way that intrinsic properties can be represented and misrepresented.

---

[31] See Thau's *Consciousness and Cognition* (New York: Oxford, 2002).

[32] Tye rejects the idea that perception is individuated by objects and their properties—*Consciousness, Color, and Content* (Cambridge: MIT, 2000), and earlier in *Ten Problems of Consciousness* (Cambridge: MIT, 1995). He argues that perception is individuated by its functional role. In the context of this paper, I am taking for granted that perceptual content is individuated by the properties and objects that the perception is of. This thesis (and thus my argument) is compatible with both externalism and internalism about perceptual content. Furthermore, the thesis is compatible with perception also being individuated by the phenomenal character of the experience.

In contrast to situation-dependent properties, mind-dependent properties cannot strictly speaking be misrepresented. No doubt, one can say that how things appear or seem to a subject *S* is accurate if, and only if, things are as they appear. So if it appears to *S* that there is something green before *S*, *S's* experience is accurate if, and only if, there is before *S* a physical object with a green surface. But that is not to deny that it is true that things appear to *S* in that way.[33] In other words, how things appear to *S* is not the kind of thing that *S* can be wrong about.

A second advantage is that it makes it possible to accommodate the fact that there is a wide range of viewing conditions or situational features that can count as normal.[34] Situation-dependent properties change as the situational features change, while intrinsic properties remain constant.

In the rest of the paper, I will elaborate on two further advantages. But first it is important to bring out one important consequence of the situation-dependency thesis. If one recognizes situation-dependent properties, then many cases of perceptual experiences that in the philosophical literature are typically categorized as illusory or as misrepresentations will turn out to be accurate perceptions—at least with regard to their representations of situation-dependent properties Take the stick partially immersed in water Given the situational features the stick appears to be bent Due to the different refraction index of water and air, the stick is presented as bent The stick is not bent, but it is accurate to perceive it to be presented as bent given the situational features

---

[33] It should be noted that D. M. Armstrong and others have argued that one can be wrong about how things appear to one—see his *A Materialist Theory of the Mind* (New York: Routledge 1968). For a discussion of the problem of misrepresenting appearance properties, see Egan, p. 16 and also p. 22, as well as Allen (manuscript). It would lead too far afield to discuss introspective fallibility and the possibility of misrepresenting appearance properties here. I will save it for another occasion.

[34] For discussion, see Hardin. Among many others, Wilfird Sellars's account of perception relies on the idea of normal viewing conditions. See his *Empiricism and the Philosophy of Mind*, with an introduction by Richard Rorty and a Study Guide by Robert Brandom (Cambridge: Harvard 1977).

Some illusions will retain their status, for instance, the Muller-Lyer illusion and similar psychological illusions The reason why one perceives one stick as longer than the other despite the fact that they are the same length is not due to representing two different situation-dependent properties The reason for the illusion is rather the Gestalt principles of convergence and divergence the lines at the sides lead the eye either inward or outward thereby creating a false impression of length

This points to an important limit of the situation-dependency thesis I take it that only basic intrinsic properties such as shapes, colors, sounds, and surface structures[35] have correlating situation-dependent properties The same is not true of high-level perceivable properties such as being happy, beautiful, or dangerous This disanalogy does not, however, imply that the situation-dependency thesis is restricted to perception of basic properties since, arguably, perception of high-level properties is dependent on perception of basic properties If this is right, then every event of perception will include perception of basic properties As a consequence, the situation-dependency thesis holds for all events of perception

III THE ARGUMENT FOR THE EPISTEMIC DEPENDENCE THESIS

I have argued that the ways objects are presented can be understood as external, mind-independent, albeit situation-dependent properties of objects Only if perceptual knowledge of intrinsic properties can be shown to be epistemically dependent on representations of the relevant situation-dependent properties, can the suggested view, however, be any better than naive direct realism Call the thesis that perceptual knowledge of intrinsic properties is epistemically dependent on representations of the relevant situation-dependent properties the *epistemic dependence thesis* The intuition on which this thesis is grounded is that one perceives an object's intrinsic properties precisely because of the way the object is presented One perceives a table to remain constant in size as one moves away from it precisely because it fills out less of one's visual field And one perceives the rectangular shape of the table precisely because of the way the shape is presented in one's egocentric frame of reference Similarly, one perceives a uniformly white wall to be uniformly white precisely because it is presented the way white walls are presented in the very lighting conditions under which one is perceiving it

---

[35] My argument has focused on the case of visual perception, but I take it to extend to all sensory modalities It would lead too far afield to go into the specifics of all sensory modalities here The argument only extends to taste and olfaction under particular assumptions of what the objects are of these two sensory modalities Thanks to Rachael Brown and Kim Sterelny for discussions on olfaction

There are many kinds of dependency relations No one denies that if a subject *S* sees a material object *o*, then *o* causally affects *S* Arguably, what is causally relevant is what I call the situation-dependent properties of the object Naïve direct realists can acknowledge such a causal dependence But by arguing that it plays a role only on a subpersonal level, they can deny that it is of any significance for a philosophical account of perceptual knowledge

Regardless of how one stands on the question of a causal-dependence, there is, I will argue, an *epistemic dependence* of perceptual knowledge of intrinsic properties on representations of situation-dependent properties The epistemic priority of the ways objects are presented given the situational features is due to the spatiotemporal nature of perception Subjects differ dramatically in the way their sensory organs are spatially extended One can imagine subjects who can perceive an object from several angles simultaneously The way the sensory organs are spatially extended depends on the particular kind of perceiver But it is a necessary feature of perception that subjects perceive from somewhere at some time If one necessarily perceives objects from a particular location at a particular time, then the way an object is presented depends on the situational features As I will show, it follows from this that representations of situation-dependent properties are the basis for knowledge of intrinsic properties Another way of expressing the same point is that they are what James Pryor calls "justification-making conditions" for perceptual knowledge of intrinsic properties[36] Consider again Peacocke's trees The nearer tree is presented as larger than the tree that is further away This difference in size with regard to their situation-dependent properties is the basis for her perceptual knowledge that the trees are the same size The subject has experiential evidence that the two trees are the same size This evidence is, however, parasitic on her evidence that the nearer tree is presented as larger than the tree that is further away from her Both layers of evidence are liable to defeat However, if evidence for the situation-dependent properties is defeated, the subject's evidence for the intrinsic properties is defeated, but not vice versa Defeaters can be understood in two ways While *undercutting defeaters* block the line of evidence from which the warrant actually arises, *rebutting defeaters* provide independent lines of evidence warranting the contrary conclusion If the subject's evidence for the situation-dependent properties is defeated, then her evidence for the intrinsic properties is undercut (and not just rebutted) It follows from this that the asymmetry of defeat is grounded in an asymmetry of warrant It is because the evidence for the situation-dependent property is in the line of evidence for the intrinsic property that defeat of the former entails defeat of the latter And it is because the evidence for the intrinsic property is not in the line of evidence for the situation-dependent property that defeat of the former does not entail defeat of the latter Thus, evidence for intrinsic properties is dependent on evidence for situation-dependent properties both with regard to defeat and warrant

---

[36] Pryor, "There Is Immediate Justification," in Ernest Sosa and Matthias Steup, eds , *Contemporary Debates in Epistemology* (Malden, MA Blackwell, 2005), pp 181–201, especially, p 182

This relation of epistemic dependence implies that perceptual knowledge of intrinsic properties is *mediated* by representations of situation-dependent properties It need not imply, however, that subjects arrive at perceptual knowledge of intrinsic properties by a conscious *inference*[37] or that intrinsic properties are not perceptually available The whiteness of the wall to my right and the roundness of the rim of the cup before me are perceptually available to me [38]

Furthermore, the thesis that there is an epistemic primacy to situation-dependent properties does not imply that they are primary in perceptual consciousness One could deny that when perceiving an object one is at all aware of how it is presented given the situational features perceptual awareness is simply of the intrinsic properties Intuitions diverge strongly on whether perceptual awareness includes the shadow play on the wall over and above the whiteness of the wall Certainly when perceiving a white wall humans typically do not attend to the details of the shadows cast on the wall But sometimes one does attend to the shadows, for instance, when one takes a realistic painter's point of view The important point for the present discussion is that even if one's attention is directed at an object's intrinsic properties, one can be aware of the way the object is presented

---

[37] For discussion, see McDowell, "Criteria, Defeasibility and Knowledge," *Proceedings of the British Academy*, LXVIII (1982) 455–79

[38] I argue for a way of understanding how a subject moves from representations of situation-dependent spatial properties to representations of intrinsic spatial properties in my "Action and Self-Location in Perception "

Certainly, when one's location in relation to an object changes, one is typically aware of something changing regarding the shape or size of the object The argument for the situation-dependency of perception requires that it is possible to be aware of the situation-dependent properties of an object without attending to them But it does not depend on how the details of the relation between awareness and attention are resolved There are many ways of understanding the relation between awareness and attention [39] One that seems particularly conducive to the present discussion is Bill Brewer's distinction between foreground and background awareness [40]

The fact that intrinsic rather than situation-dependent properties typically occupy the attention of human perceivers may appear to stand in conflict with the thesis that there is an epistemic primacy to situation-dependent properties But it is important to distinguish between what is primary in perceptual consciousness from what is causally primary and what is epistemically primary I am arguing only that situation-dependent properties are epistemically primary Epistemic primacy can but need not be understood to entail causal primacy [41] Regardless of how one stands on the relation between epistemic and causal primacy, no implication about phenomenological primacy follows from either epistemic or causal primacy The epistemic dependency thesis does not imply that situation-dependent properties are in the foreground of a subject's perceptual awareness, although it is of course possible that they are The thesis is even compatible with the claim that subjects are not aware of situation-dependent properties, at least if one accepts that for a subject to have evidence, the subject need not be aware of this evidence [42] If a subject can have evidence without being aware of this evidence, then the fact that intrinsic properties typically occupy the attention of at least human perceivers does not stand in conflict with the thesis that perceptual knowledge of the whiteness of the wall is epistemically dependent on representing the way it is presented given the lighting conditions. I take it to be plausible that subjects are always to some extent aware of situation-dependent properties, but my thesis does not depend on it.[43]

---

[39] For helpful accounts of this distinction, see Campbell, *Reference and Consciousness* (New York Oxford, 2002), and Ned Block, "Mental Paint," in Martin Hahn and Bjørn Ramberg, eds , *Reflections and Replies Essays on the Philosophy of Tyler Burge* (Cambridge MIT, 2003), pp 165–200

[40] See Brewer, *Perception and Reason* (New York Oxford, 1999), especially sections 5 3 1 and 7 4 3

[41] For a defense of the idea that epistemic dependence entails causal dependence, see Hilary Kornblith, "Naturalizing Rationality," in Newton Garver and Peter H Hare, eds , *Naturalism and Rationality* (Buffalo, NY Prometheus, 1986), pp 115–33 For a critical discussion of the thesis, see Robert Audi, *The Structure of Justification* (New York Cambridge, 1993), pp 99ff

[42] Michael Bergmann argues that a subject's belief can be justified by something even if the subject is not aware of that thing See his *Justification without Awareness* (New York Oxford, 2006), especially chapter 1

By pulling apart what is epistemically primary from what is phenomenologically primary the suggested view differs from both standard direct and indirect realist views. On both views, the properties and objects that are epistemologically primary are also phenomenologically primary. According to naive direct realists intrinsic properties are argued to be both epistemologically and phenomenologically prior. According to indirect realism, what is directly present to the mind is argued to be a mind-dependent object or property (that correlates to what I call situation-dependent properties). Arguably, indirect realists are committed to the thesis that these mind-dependent objects or properties are both phenomenologically and epistemologically prior. Because on the suggested view there are two kinds of external, mind-independent properties, what is epistemologically prior need not coincide with what is phenomenologically prior.

Now, it is empirically well established that the process of bringing about representations of situation-dependent and intrinsic properties are mutually dependent in the human visual system insofar as information about situation-dependent properties is used to gain information about intrinsic properties and vice versa.[44] Acknowledging this fact is compatible with accepting the epistemic dependency thesis insofar as the epistemic dependence between the two representations is independent of the process by which these representations are brought about. The epistemic dependence is not a matter of how the visual system arrives at the representations but a matter of the relations of defeasibility and warrant.

I have argued that perceptual knowledge of intrinsic properties is epistemically dependent on representing situation-dependent properties It is important to note that the epistemic dependency thesis does not require that the representations of situation-dependent properties constitute knowledge The thesis is weaker representations of situation-dependent properties are the *basis* for knowledge of intrinsic properties So I am presupposing that evidence need not constitute knowledge Moreover, I take it to be plausible that representations of situation-dependent properties need not have a propositional structure My thesis is, however, neutral on this issue If, however, representations of situation-dependent properties do *not* constitute propositions, then the epistemic dependency thesis can only be true if evidence for knowledge does not require that this evidence is propositional [45]

---

[43] It has been argued that concepts such as "is red" should be analyzed in terms of concepts used to pick out sense-experiences, such as "looks red." The question of whether representations of intrinsic properties are epistemically dependent on representations of situation-dependent properties is orthogonal to the question of any semantic dependence of "material-object" predicates on "looks"-predicates. For different versions of this thesis, see Chisholm, and see Roderick Firth, "Coherence, Certainty, and Epistemic Priority," this JOURNAL, LXI, 19 (October 1964): 545–57. For a critical discussion, see Sellars, "Givenness and Explanatory Coherence," this JOURNAL, LXX, 18 (October 1973): 612–24.

[44] Thanks to Kim Sterelny for pressing me on this point. For an excellent discussion of the processing relation between viewpoint-invariant and viewpoint-relative representations, see Jesse Prinz, *Furnishing the Mind* (Cambridge: MIT, 2002), in particular chapters 6 and 7.

I have argued that perceptual knowledge of the intrinsic properties of an object is epistemically dependent on representing the relevant situation-dependent properties If representing intrinsic properties is dependent on representing their situation-dependent properties, then the representation of situation-dependent properties must be a necessary part of perceptual content So perceptual content will crucially differ from the content of a thought about the same object insofar as a thought about an object does not necessarily represent the object's situation-dependent properties [46]

There are many other ways of thinking about the difference between perceptual content and the content of thought One is to say that the fundamental difference between perception and thought is that there are phenomenal or sensational properties present in perception but not in thought [47] Another is to say that the difference between the content of perception and the content of thought is a structural or functional difference Robert Brandom argues, for instance, that the role of thoughts in language is inferential, while the role of perceptions is noninferential [48] According to McDowell,[49] we employ the very same conceptual capacities passively in experience that we employ actively in thoughts Fred Dretske distinguishes the content of experience and thought with regard to the origin of their functions [50] According to Tye,[51] thoughts are representations in a linguistic medium, while experiences are representations in an image-like medium I stay agnostic on whether perception and thought differ phenomenally and whether their content differs structurally or functionally over and above the difference in what is represented If the argument I have given holds at all, it holds regardless of what stance one takes on these other possible differences

---

[45] It is typically taken for granted that evidence need not constitute knowledge and need not be propositional—recently, for instance, by Bergmann Williamson argues that all evidence is knowledge and that all evidence is propositional—*Knowledge and Its Limits* (New York Oxford, 2000) It would go well beyond the scope of this paper to consider Williamson's argument here

[46] Demonstrative thoughts are a special case Arguably they are based on perception and to that extent mirror the situation-dependency of perception

[47] There are exceptions Terence Horgan and John Tienson argue that mental states, such as beliefs and desires, have phenomenal character as well, See their "The Intentionality of Phenomenology and the Phenomenology of Intentionality," in Chalmers, ed , *Philosophy of Mind* (New York Oxford, 2002), pp 520-33

[48] See Brandom's *Making It Explicit* (Cambridge Harvard, 1994), chapter 4

[49] See McDowell's *Mind and World*

IV THE ARGUMENT FOR THE DIRECT PERCEPTION THESIS

Due to the role of appearances in the tradition of Locke, Berkeley, and Hume, the idea that objects are presented a certain way has been considered to be at odds with the intuition that we directly perceive objects The notion of direct perception has been understood in many different ways As I understand it, one perceives an object directly if one does not perceive it in virtue of perceiving another thing This conception must be contrasted with a notion according to which to perceive something directly would be to be perceptually aware of it without bringing any capacities to bear on what is given in perception that are not narrowly construed sensory capacities [52]

If the ways objects are presented given the situational features are distinguished both from how they appear to be and from representations of the ways objects are presented, then there is no conflict between the intuition that we perceive objects directly and the claim that objects are presented a certain way given the situational features Since situation-dependent properties are actual properties of an object, we perceive an object directly by perceiving its situation-dependent properties directly

The objection could be raised that this way of thinking about perception is committed to an epistemic intermediary and thus subject to the very same objections that can be raised against indirect realism According to indirect realism, one perceives objects indirectly by having a special kind of mind-dependent object present to mind. Arguably, such a view introduces an epistemic intermediary between one's perception and the physical objects that one's perception is of, thereby cutting us off from reality. One could say that if one perceives the intrinsic properties of objects in virtue of perceiving their situation-dependent properties, then one is similarly cut off from the object as it is.

---

[50] See Dretske's *Naturalizing the Mind* (Cambridge MIT, 1995)

[51] See Tye's *Consciousness, Color, and Content*

[52] The first understanding has been considered a non-epistemological characterization of direct perception, the second an epistemological characterization of direct perception (see, for instance, Jackson, *Perception*) There are other ways of characterizing direct perception epistemologically, the differences depending largely on what one means by epistemological Paul Snowdon characterizes an epistemological understanding of direct perception as the idea that something is directly perceived if and only if it is noninferentially known—see his "How to Interpret 'Direct Perception'," in Tim Crane, ed , *The Contents of Experience* (New York Cambridge, 1992), pp 48–78

In contrast to the mind-dependent items that the indirect realist invokes, situation-dependent properties do not constitute an object that is distinct from the physical object in view. The situation-dependency thesis does not involve a commitment to any object over and above the physical object that one's perception is of. An object's intrinsic properties and its situation-dependent properties are properties of the very same object. Therefore, acknowledging situation-dependent properties makes it possible to account for the ways objects are presented while maintaining that subjects perceive objects directly. Subjects perceive objects directly in virtue of perceiving their situation-dependent properties directly. So the suggested way of thinking about perception differs fundamentally from indirect realist views. While the indirect realist argues that we perceive physical objects indirectly by having something mind-dependent present to mind, the thesis presented here is that we perceive objects directly by perceiving their situation-dependent properties directly.

This will not impress a skeptic. But the point of understanding how objects are presented as external and mind-independent is not to fend off skeptical worries. The point is rather to show that much of what is typically thought of in terms of the subjectivity of perception should rightly be thought of in terms of the situation-dependency of perception. Skeptical worries arise for the situation-dependent properties in just the way that they arise for the existence of an external world, precisely because they are elements of the external world.

V. CONCLUSION

My aim was to show that in order to explain how perception can give us knowledge of objects, the way one gains knowledge through perception must be taken seriously. The way one gains knowledge through perception is by being presented with objects given situational features, such as the lighting conditions and one's location in relation to the object. If this is right, then perception is necessarily situation-dependent. I spelled out this situation-dependency thesis by arguing that the ways objects are presented can be understood in terms of situation-dependent properties: external, mind-independent properties that an object has given the situational features

I argued that recognizing situation-dependent properties yields four advantages First, it makes it possible to acknowledge that perceptual experiences are not just individuated by the objects they are of, but by the ways those objects are presented given the situational features On standard views, the conditions an experience must satisfy to be accurate is only to attribute the right intrinsic properties to the right object the way the object is presented does not matter for whether the experience is accurate In contrast to such views, I have argued that the ways objects are presented are the kind of things that can be represented and misrepresented in just the way that intrinsic properties can be represented and misrepresented By understanding the ways objects are presented in terms of mind-dependent, situation-dependent properties we can recognize that more of experience is subject to accuracy conditions than standard views acknowledge The conditions an experience must satisfy to be accurate is not just to get the intrinsic properties right, but is also to get the situation-dependent properties right So the ways objects are presented are objective contents of subjective acts of consciousness in the very same way as objects and their intrinsic properties are objective contents of subjective acts of consciousness Insofar as situation-dependent properties can be represented and misrepresented in the very same way that intrinsic properties can be represented and misrepresented, they are systematically related to the content of states that ground our beliefs about the world

Second, recognizing situation-dependent properties allows for a way to accommodate the fact that there is a wide range of viewing conditions or situational features that can count as normal Situation-dependent properties change as the situational features change, while intrinsic properties remain constant

Third, I argued that perceiving the intrinsic properties of objects is epistemically dependent on representing their situation-dependent properties If perceiving the intrinsic properties of objects is epistemically dependent on representing their situation-dependent properties, then representations of situation-dependent properties are a necessary part of perceptual content Insofar as perceptual content necessarily involves representations of situation-dependent properties, it crucially differs from the content of a thought about the same object

Finally, recognizing situation-dependent properties makes it possible to embrace the motivations that lead to phenomenalism and indirect realism by recognizing that objects are presented a certain way, while holding on to the intuition that subjects directly perceive objects The suggested way of thinking about perception is

UNIVERSITY CALCUTTA CAMPUS

similar to an indirect realist approach in that perceiving the intrinsic properties of objects is dependent on perceiving something else In contrast to an indirect realist approach the intermediary is, however, not a subjective, mind-dependent intermediary The intermediary is rather a second kind of external, mind-independent property of the very same object, namely a situation-dependent property Perception of intrinsic properties is dependent on perception of situation-dependent properties

Views according to which perception involves representations of mind-dependent properties (or objects) can recognize situation-dependent properties If they do, there will be mind-dependent properties (or objects) that stand in representation relations both to intrinsic properties and to situation-dependent properties But although formally consistent with such views, recognizing situation-dependent properties undermines at least one central motivation that has traditionally led to such views If one recognizes situation-dependent properties, no appeal to mind-dependent properties is necessary to explain how it can be that there is a way that objects look that is not accounted for by representing their intrinsic properties If this is the reason for introducing mind-dependent properties (or objects), then one might as well drop them once situation-dependent properties are acknowledged

SUSANNA SCHELLENBERG

Australian National University

IS REALISM REALLY BAD FOR YOU? A REALISTIC RESPONSE*

Realism about oneself and one's circumstances has long been regarded as a hallmark of mental health and authentic happiness by philosophers and psychologists It has also long invited skepticism from some quarters Recently, this skepticism has found new support in the work of some social psychologists, who claim that far from being *essential* for mental health or happiness, realism can be *bad for you* Certain positive illusions about yourself, they say, are more conducive to health and happiness than a high degree of realism I will argue that, properly understood, realism really is good for you I will begin by showing the importance of realism in psychological, philosophical, and everyday thought about health and happiness

Most psychologists and counselors regard healthy individuals as reality-oriented, as "able to take in matters" they wish were different, "without distorting them to fit these wishes "[1] In a Socratic statement, Gordon Allport declares that "an impartial and objective attitude toward oneself is a primary virtue, basic to the development of all others And so if any trait of personality is intrinsically desirable, it is the disposition to see oneself in perspective "[2]

Abraham Maslow and Carl Rogers also echo the idea that realism is central to mental health and happiness, where realism involves not only characteristically facing up to the facts, but also evaluating them by realistic standards and responding in realistic ways In the words of Paul B Baltes and his colleagues, realistic people have "rich factual knowledge about human nature" and "rich procedural knowledge

---

* I am grateful to Ronald de Sousa, Erica Roedder, Dan Haybron, Anna Alexandrova, Linda Zagzebski, Chris Swoyer, Joel Kupperman, and the editors for their helpful comments on earlier versions I am also grateful to Candace Upton and the Carl M Williams Institute of Ethics and Values, the University of Missouri at St Louis, and St Louis University for the opportunity to present this paper and benefit from the discussions

[1] Marie Jahoda, "The Meaning of Psychological Health," *Social Casework*, XXXIV (1953) 349, cited in C Randall Colvin and Jack Block, "Do Positive Illusions Foster Mental Health? An Examination of the Taylor and Brown Formulation," *Psychological Bulletin*, CXVI, 1 (July 1994) 3–20

[2] Gordon W Allport, *Personality A Psychological Interpretation* (New York Holt, Rinehart, and Winston, 1937), p 422, cited in Colvin, Block, and David C Funder, "Overly Positive Self-Evaluations and Personality Negative Implications for Mental Health," *Journal of Personality and Social Psychology*, LXVIII, 6 (1995) 1152–62

0022-362X/08/0502/85–107

about ways of dealing with life problems."[3] Such people are also strongly disposed to act accordingly. Realism understood thus is the chief part of practical wisdom, and includes both a certain *attitude*—reality-orientation—and a certain *achievement*. Realism does not, however, entail perfection in either of these dimensions. Indeed, it is because realistic people recognize their (inevitable) lack of perfection that they are, as Maslow puts it, self-actualizing, that is, inclined to constantly strive for growth while accepting their own unchangeable weaknesses as individuals and as human beings "without chagrin or complaint."[4] Such people are also happy, because they more fully integrate the "pleasure principle" with the "reality principle," and more fully attain values or virtues such as "serenity, kindness, courage, honesty, love, unselfishness."[5] On this picture, realism is both instrumental to happiness, and partly constitutive of it.

Likewise, on Rogers's view, the healthy or fully functioning individual is open to experience, distorting neither his perceptions of the world to fit his conception of himself, nor his conception of himself to fit his perceptions of the world.[6] He is self- rather than other-directed, avoids façade, pretense, and defensiveness, and is open to inner and outer reality.[7] In virtue terms, the fully functioning individual is self-directed or autonomous,[8] honest with himself, and honest in the way he presents himself to others. He also has a sense of pride and self-worth that puts a negative value on pleasing others as an ultimate goal (*op. cit.*, p. 182). On Rogers's view, it is only a fully functioning individual who leads the good life—a rich, full, challenging, and exciting life of continual growth—for only such an individual has the courage to launch himself "fully into the stream of life."[9]

Maslow's and Rogers's conceptions of the good life as requiring good functioning, and good functioning as requiring realism and an active life of continual growth with honesty, courage, and autonomy, has much in common with ancient ideas of happiness (*eudaimonia*) or the good life Thus, Aristotle tells us that the well-functioning and happy individual must be virtuous, and that a virtuous individual must have practical wisdom (*phronesis*), that is, a high degree of knowledge of himself, of the things that matter in human life and of how to achieve them, and the emotional and intellectual dispositions to feel, decide, and act accordingly[10] Realism is, thus, part of Aristotle's very conception of practical wisdom, and the wise individual's virtues include autonomy (the ability to direct himself by his own practical reason), honesty with himself, and honesty in his self-presentations[11] Aristotle argues that only such an individual can achieve *eudaimonia* or happiness In contemporary terms, a *eudaimonic* life is both objectively worthwhile and subjectively fulfilling, that is, enjoyable and meaningful from the individual's own point of view[12] Neither subjective fulfillment nor objective worth alone is enough for happiness understood as *eudaimonia* To avoid confusion, I will henceforth reserve the terms 'happiness' or 'true/genuine happiness' to mean *eudaimonia*,[13] 'subjective happiness' to mean "subjective fulfillment" or "sense of happiness," and 'unhappiness' to mean subjective unhappiness, lack of worth, or both

---

[3] Paul B. Baltes, Ute Kunzmann, and Antje Stange, "Research Project 6: Wisdom: The Integration of Mind and Virtue," Center for Lifespan Psychology, http://www.mpib-berlin.mpg.de/en/forschung/lip/pdfs/research_project_6.pdf, pp. 196–200.

[4] Maslow, "Self-Actualizing People: A Study of Psychological Health," in R. J. Lowry, ed., *Dominance, Self-Esteem, Self-Actualization: Germinal Papers of A. H. Maslow* (Monterey, CA: Brooks/Cole, 1973), pp. 177–98, at p. 183.

[5] Maslow, *Toward a Psychology of Being* (New York: Wiley, 1999, 3rd ed.), p. 171.

[6] Rogers, "A Therapist's View of the Good Life: The Fully Functioning Person," in H. Howard Kirschenbaum and V. L. Henderson, eds., *The Carl Rogers Reader* (New York: Houghton Mifflin, 1989), pp. 409–29. Rogers makes it clear that his fully functioning person is not the perfectly functioning person, but simply the person who is constantly moving towards better and better functioning (pp. 410–11, p. 416).

[7] Rogers, "The Person in Process," in *The Carl Rogers Reader*, pp. 155–97, at p. 182.

[8] There are, of course, many conceptions of autonomy, but self-directedness is at the heart of all of them.

[9] Rogers, "A Therapist's View of the Good Life," in *The Carl Rogers Reader*, p. 420.

The ideal of a happy life as one that is both subjectively fulfilling and objectively worthwhile and, hence, free of illusion or delusion, is also widespread among ordinary people Thus, it is important to most of us that we actually *achieve* our goals, and not just that we have the pleasurable but illusory *experience* of achieving them, that the people we care about *actually* fare well, and not just that we deludedly *believe* that they do This is why, when presented with Robert Nozick's experience machine thought experiment, students overwhelmingly reject the illusory life of perfect bliss on the machine for the reality of imperfect happiness off the machine [14] The importance we place on being in touch with our lives also explains why it makes sense to think of a man who is confident in the loyalty of his unfaithful friends and the excellence of his mediocre work as an "unhappy man " Regardless of his sense of fulfillment, he is "unhappy" because he is living in a "Fool's Paradise "

---

[10] Aristotle, *Nicomachean Ethics* (*NE*), II 7, VI 12, 13 Both for Aristotle and for us, self-knowledge includes knowledge of one's character, as well as of one's interests, abilities, and temperament It also includes knowledge of the empirical conditions of one's life Again, both for Aristotle and for us, the things that matter in human life include the things and activities needed for survival, human relations of love, work, and play, and intellectual and creative activity The main difference between Aristotle and us is that we also recognize physical labor and productive economic activity as important exercises of our human capacities

[11] The honest or truthful person, says Aristotle, is "truthful [about himself] both in what he says and in how he lives simply because that is his character" (*NE*, IV 7)

[12] The enjoyment and sense of meaning come both from virtuous activity, which is necessarily pleasurable and meaningful to the virtuous individual, and from certain important external goods, such as friends or honor (Aristotle, *NE*, I 8, IV 3, and VIII 1)

[13] Some readers might object to using the word "happiness" thus, but the practice is now quite widespread among both philosophers and psychologists See, for example, Martin Seligman, "Eudaemonia, The Good Life A Talk with Martin Seligman," http //www edge org/3rd_culture/seligman04/seligman_index html, Julia Annas, *The Morality of Happiness* (New York Oxford, 1993), and Nicholas White, *A Brief History of Happiness* (Malden, MA Blackwell, 2006) Moreover, in wishing the best to their children, people wish them "happiness," distinguishing happiness as the highest good from happiness as subjective fulfillment by using epithets like "true" or "genuine" happiness for the former

The desire to be in touch with reality extends to our aspirations for who we want to be It is important to many, if not most, of us that we actually *live* worthwhile lives, and not simply *believe* that we do We want to live lives that are worthwhile by some objective standard, a standard that applies to creatures with our needs, interests, and capacities—including the capacity for asking what sort of life counts as worthwhile In wishing for the good life, many of us wish for a life in which, as Susan Wolf puts it, subjective meaning meets objective worth [15] This is why we tend to agree with Aristotle that a life of passive pleasures is fit only "for grazing animals" (*NE*, I 5) [16]

Realism, then, has been widely seen as both instrumentally and conceptually necessary for happiness understood as subjective fulfillment in an objectively worthy life Realism is an important means to subjective fulfillment, and both a means to, and conceptually necessary for, objective worth The happy individual is reality-oriented and informed about the important facts of her own life and human life in general, characteristically evaluates and responds to events in the light of these facts, and has a sense of fulfillment that is grounded in them

---

[14] Nozick discusses this thought experiment in *Anarchy, State, and Utopia* (New York Basic Books, 1974), pp 42–45, and in *The Examined Life* (New York Simon and Schuster, 1989), pp 104–08 I have repeated it with students semester after semester No matter how attractive I make the experience machine, I can never persuade more than 5% of undergraduates or 10% of graduate students to sign up for it (I note the difference in percentages without comment)

[15] Susan Wolf, "Happiness and Meaning Two Aspects of the Good Life," *Social Philosophy and Policy*, XIV, 1 (Winter 1997) 207–25 Wolf argues that happiness and meaning are both part of the good life Since by "happiness" she means what I call "enjoyment of life" and by "good life" what I call "a life of happiness as the highest good" or *eudaimonia*, her conception of the highest good is essentially the same as mine

[16] As I have argued in "Experiments in Living" (*The Philosophers' Magazine*, XXXV (3rd quarter 2006) 58–61), John Stuart Mill's distinction between higher and lower pleasures is also based on the idea that some pleasures are more *worthy* of a human being, befitting the sense of dignity that Mill believes all human beings possess to a greater or lesser degree

This picture of reality-based happiness and mental health is not only internally coherent, it is also intuitively plausible and highly attractive Directly or indirectly, however, it is this picture of realism that skeptics challenge The challenge seems to come both from common sense, and from social and cognitive psychology, with the latter providing support for the former The challenge is empirical in nature, directed at the claim that realism is an important means to subjective happiness and mental health But if the challenge is successful, it also undermines, as we shall see, the normative and conceptual claim that realism is partly constitutive of mental health and objective worth and, thus, of (*eudaimonic*) happiness This makes it even more important to examine the skeptics' criticisms and see if they are well founded I will argue that they are not In the next section I will address the commonsense challenge, and in section II the social scientific challenge

I REALISM WITHIN THE LIMITS OF HUMAN NATURE ALONE

The claim that being realistic (reality-oriented and informed about the important facts of your own life and human life in general, and disposed to act accordingly) is necessary for true happiness seems to fly in the face of an obvious truth, namely, that it is possible to know too much for your own good The discovery that your kindly grandfather was a gun for hire, or that your loving husband, now deceased, was having an affair with your best friend, seem to be discoveries best not made Moreover, a reality-oriented person—someone who is disposed to stay in touch with the important facts of her life—is far more likely to make such discoveries than someone who is willing to let sleeping dogs lie Such a person is also more likely to be aware of her own flaws and the flaws of the human species, an awareness not exactly calculated to boost the mood This seems to show that being reality-oriented and informed is inimical to happiness, that, as *Ecclesiastes* laments, "he that increaseth knowledge increaseth sorrow "

If *Ecclesiastes* is right, then there is a tension at the heart of the idea that happiness is subjective fulfillment in an objectively worthwhile and, thus, realistic, life For if knowledge generally increases sorrow, we end up with the strange result that to the extent we are *unrealistic*, our lives are, by hypothesis, lacking in worth, whereas to the extent we are *realistic*, we are likely to feel unfulfilled We can try to make our lives *either* objectively worthwhile *or* subjectively fulfilling, but we cannot expect to make them both

But *Ecclesiastes* exaggerates. Although knowledge of some facts *can* lead to a sorrow so great that it leads to a decrease of overall happiness, the sorrow often comes not from knowledge of the truth as such, but from the untimeliness of the discovery. Like surgical procedures, unpleasant truths may need a long prep time. Recognizing this and preparing oneself psychologically for a potentially devastating revelation before proceeding with one's investigations is the better part of realism—of wisdom about one's human and individual limitations. And wisdom, after all, is what genuine realism is about, not a mechanical harvesting of all important truths, with nary a care for the how, when, or what for.

Nevertheless, it must be granted that there are some truths that are both important and too terrible to face ever. Learning the brutal circumstances of your only child's death might lead to a grief so unbearable that it can only diminish your happiness. And again, the reality-oriented person is more likely to learn such truths than someone who habitually avoids unpleasant facts.

This example, however, is less troublesome than at first it appears to be. What it shows is that being reality-oriented *can* lead to a discovery that diminishes our happiness by increasing our grief and adding nothing of worth to our lives. It does not show that being reality-oriented is *more* likely to diminish our happiness than its contrary, being oriented to self-deception.[17] On the contrary, being characteristically self-deceived is a far more reliable route to grief, since reality is impervious to our wishes, and unwelcome facts do not obligingly retreat from the scene when we choose to ignore them. Thus, whereas the self-deceptive are likely to skin their shins on the rough edges of reality, the reality-oriented, acknowledging that life abounds with pitfalls, prepare themselves to deal with them.

All the same, my reply covertly admits that when the facts are devastating, we might be better off remaining ignorant of them. This seems to contradict my view that genuine happiness requires being informed about, and responding appropriately to, important truths about our lives. But I am not contradicting this claim. What I am claiming is that, when knowledge of a particular fact can only crush us, whereas ignorance of it has no harmful consequences, then ignorance, even illusion, beats knowledge. Some important truths, like some grave dangers or physical pains, are beyond our capacity to bear far from providing an opportunity to make a courageous or ennobling response, they crush us under their weight In the example at hand, when you recover enough from your grief to consider yourself somewhat happy again, some of your positive feelings and evaluations will be made possible only by your ignorance of the circumstances of your child's death But my theory can agree with common sense that some subjective happiness based on ignorance or illusions is better than total misery based on true beliefs [18]

---

[17] Mike W. Martin's definition of purposeful self-deception as an evasion of full acceptance and integration of some fact, or supposed fact, usually painful or difficult, with one's other beliefs, values, and emotions, perhaps best captures the main elements of self-deception identified by Sartre, Herbert Fingarette, and others (Martin, *Self-Deception and Morality* (Lawrence: Kansas UP, 1986), pp. 14, 15).

Suppose, however, that the brutal death of your child is the result, in part, of your own habitual negligence Does my theory still hold that ignorance is better than knowledge? The answer to this question depends on the answer to the question why genuine happiness requires knowledge of one's character and actions As I have been arguing, such knowledge is important for improving oneself and avoiding negligent acts in the future, as well as for making a well-grounded evaluation of one's life But if you are already aware of your habitual negligence, have taken it into account in evaluating your life, and are working to improve yourself, then all that knowledge of this particular act can do is make your evaluation of your life more accurate while adding guilt and grief to grief So, knowledge of this act is better for you only if nothing else can shake you out of your habitual negligence and prevent future tragedies If this is the case, *and* you use your new-found self-knowledge to grow in wisdom and virtue, then, and only then, on my view, are you better off knowing about your role in your child's death But this is no longer counterintuitive For even as the knowledge increases your subjective unhappiness, it both protects you from further disaster (and, thus, from further unhappiness), and contributes to the objective worth of your life Thus it contributes to your overall happiness or, at least, mitigates your overall unhappiness

The lesson of these two examples is that although being realistic (reality-oriented and informed about the important facts of your own life and human life in general, and disposed to act accordingly) is an important means to subjective fulfillment, and both an important means to, and conceptually necessary for, objective worth, it can also come into conflict with subjective fulfillment This is the chief reason why (*eudaimonic*) happiness, which requires both subjective fulfillment and objective worth, is so difficult to achieve Too many objections to the thesis that (*eudaimonic*) happiness entails realism either forget that such happiness is not simply subjective fulfillment, or assume, mistakenly, that the objective and subjective dimensions of happiness can never conflict

---

[18] The reasons I have given for this conclusion apply as well to the objection (insofar as it is true) that some people would be left in utter despair without the crutch of the illusory and irrational belief in a benevolent deity It is worth noting that, since our epistemic and moral limitations make complete knowledge of important facts and perfect reality-orientation impossible for us, it is safe to say that everyone's sense of happiness (or, for that matter, unhappiness) is, to some extent, based on ignorance and illusion

The scientific challenge to realism, however, has other arguments in its arsenal

II THE SCIENTIFIC CHALLENGE

II 1 In two influential articles, Shelley Taylor and Jonathon Brown have argued that a vast body of empirical research in social and cognitive psychology suggests that most people harbor certain mild positive illusions about themselves, and that, contrary to the conventional wisdom, these illusions are not only not opposed to mental health or happiness, they tend to promote them [19]

Before discussing the substantive issues, it is worth noting that, as is standard in contemporary psychology, Taylor and Brown regard happiness as a purely subjective state But other than suggesting that subjective happiness is identical with, or associated with, contentment, they never tell us what they mean by subjective happiness As I argued in the introduction, however, for most people subjective happiness also includes enjoyment and a sense of meaning This, as we shall see, is relevant to a proper evaluation of Taylor and Brown's claim that positive illusions promote subjective happiness [20]

To return to the substantive issues Taylor and Brown contend that people with certain "pervasive, enduring, and systematic," but mild, positive illusions about themselves and the world are generally a cheerful bunch, whereas those who are more realistic are generally somewhat depressive (IW 194) The illusions in question are "unrealistically positive self-evaluations, exaggerated perceptions of control or mastery, and unrealistic optimism" (PI 193) Taylor and Brown do not tell us to what extent these illusions are a matter of simple ignorance and to what extent a matter of self-deception or defensiveness [21] As I shall argue, however, this distinction makes quite a bit of difference to the bearing of Taylor and Brown's argument on the claim about the importance of realism in mental health

---

[19] Taylor and Brown, "Illusion and Well-Being A Social Psychological Perspective on Mental Health," *Psychological Bulletin*, ciii (1988) 193–210, at p 193 (hereafter IW) See also Taylor and Brown, "Positive Illusions and Well-Being Revisited Separating Fact from Fiction," *Psychological Bulletin*, cxvi, 1 (1994) 21–27, at p 5 (hereafter PI)

[20] There are two other problems with Taylor and Brown's understanding of subjective happiness and mental health (i) They regard subjective happiness as a measure of mental health But subjective unhappiness is a perfectly appropriate response to severe misfortune, and large numbers of people in the world, whose lives are an endless struggle for survival, are appropriately (more-or-less) unhappy (ii) Taylor and Brown equate psychological well-being with mental health But since "psychological well-being" entails subjective happiness whereas mental health does not, well-being cannot be identical with mental health

They also argue that the empirical data show that the somewhat self-deceived or ignorant but cheerful people (call them the Upbeat) do better than the depressed realists (the Downbeat) along all the traditional dimensions of mental health other than realism "contentment, positive attitudes toward the self, the ability to care for and about others, openness to new ideas and people the ability to perform creative and productive work, and the ability to grow and self-actualize, especially in response to stressful events" (PI 22) These illusions foster better adjustment as well as "better life functioning "[22]

Although Taylor and Brown do not talk about wisdom or virtue, we can infer from these claims that the Upbeat are more likely to have qualities that are generally thought of as wise and virtuous kindness, tolerance, generosity, a perspective that allows them to take things in stride, and the courage to grow and self-actualize Hence, in challenging the view that being reality-oriented and informed about oneself are instrumentally important for subjective happiness and mental health, Taylor and Brown are also, by implication, challenging the view that being reality-oriented and informed are necessary for wisdom and virtue

If Taylor and Brown are right, it follows that wisdom requires cultivating the art of tempering self-knowledge with mild self-deception, while taking care not to let it get out of hand Indeed, Taylor and

---

[21] Taylor and Brown do distinguish positive illusions from defense mechanisms, but their distinction seems like a distinction without a difference Positive illusions, they say, "are directly responsive to threatening circumstances, whereas defenses are inversely responsive to threatening information Thus, for example, advancing cancer patients typically do not deny or repress information about their deteriorating condition They are aware that their circumstances have worsened, but within the context of this acknowledgment, they may put a more optimistic spin on their circumstances than conditions warrant" (PI 25) But a more optimistic spin than is warranted by the facts *insofar as one is aware of them* (an important addition) is possible only if some fact is denied or repressed The only discernible difference between defense mechanisms and optimistic spins is that the former are extreme and the latter mild

[22] Taylor, R L Collins, L A Skokan, and L G Aspinwall, "Maintaining Positive Illusions in the Face of Negative Information Getting the Facts without Letting Them Get to You," *Journal of Social and Clinical Psychology*, vii (1989) 114–29, at pp 115–16, cited in Colvin and Block, p 4

Brown explicitly recommend nurturing the capacity for self-deception (or, as they prefer, for unwarranted "optimistic spins")

> The individual who responds to negative, ambiguous, or unsupportive feedback with a positive sense of self, a belief in personal efficacy, and an optimistic sense of the future will, we maintain, be happier, more caring, and more productive than the individual who perceives this same information accurately and integrates it into his or her view of the self, the world, and the future In this sense, *the capacity to develop and maintain positive illusions may be thought of as a valuable human resource to be nurtured and promoted,* rather than an error-prone processing system to be corrected (IW 205, italics mine)

For all the attention their thesis has generated, however, Taylor and Brown's evidence for their thesis is surprisingly weak, and their arguments are riddled with conceptual and logical problems[23] Unfortunately, too many people, including philosophers, have tended to accept their thesis as valid without subjecting it to critical scrutiny[24] Let us start by asking about the evidence and main arguments for their conclusion

II 2 The main source of evidence of widespread positive illusions in "Illusion and Well-Being" is experiments with college students, although to broaden the basis of their claims, Taylor and Brown also cite some studies of terminally ill individuals in "Positive Illusions and Well-Being Revised" (see section ii 6 below) The experiments with students show that most students rate themselves higher on their abilities, achievements, degree of control, and future prospects than they do most of their peers (IW 195–97) Sixty percent of respondents in surveys of moods also believe that "they are happier than most people "[25] Taylor and Brown give the following arguments for thinking that the students' positive self-attributions are somewhat illusory (i) it is logically impossible for most students to be happier or higher in their abilities, achievements, or degree of control than most other students at the same university, (ii) most students rate themselves more positively on their attributes than do observers, and (iii) it is logically impossible for most students to have better life prospects than most other students, moreover, (iv) base rate data cannot justify their optimism about the future (IW 2, PI 22) [26]

---

[23] A google search, accessed September 6, 2007, reports 1,688 citations to date for the 1988 article (IW), and 240 citations for the 1994 article (PI) In addition, Taylor's book, *Positive Illusions Creative Self-Deception and the Healthy Mind* (New York Basic Books, 1991) and the press about it have brought the message that illusions are good for you to thousands of people

[24] For example, in "On Overrating Oneself and Knowing It" (*Philosophical Studies,* cxxiii (2005) 115–24), Adam Elga's unquestioned starting point is that it has been shown that people are subject to "persistent and widespread positive illusions (about themselves)" (p 117) Again, citing Taylor and Brown, Hilary Kornblith states "While it was once the standard view among psychologists that being emotionally well-adjusted goes hand in hand with an accurate understanding of oneself and the world around one, the evidence against this claim is now quite strong" ("What Is It Like to Be Me?" *Australasian Journal of Philosophy,* lxxvi, 1 (1998) 48–60) I have seen only one critical philosophical article, David A Jopling's "'Take Away the Life-Lie ' Positive Illusions and Creative Self-Deception," *Philosophical Psychology,* ix, 4 (1996) 525–44 Jopling criticizes the Taylor-Brown thesis on grounds of "ecological validity and phenomenological realism," as well as on the substantive ground that "positive illusions diminish the range of reactive other-regarding attitudes and emotions" (p 525) Unfortunately, Jopling's substantive criticism is vitiated by his use of an example of a thoroughly self-deceived individual as a counterexample to Taylor and Brown's thesis of the connection between illusions and mental health, in spite of Taylor and Brown's insistence that they are defending only mild illusions The best critical article is by the social psychologists, Colvin and Block, "Do Positive Illusions Foster Mental Health?" They argue that "the logic and empirical evidence used to relate mental health to [positive illusions] failed to substantiate Taylor and Brown's thesis," and that "more recent studies on positive illusions and mental health also failed to lend support" to this thesis

It should be noted that college students are not exactly representative of the American population at large Given their youth and inexperience, we should expect students to be particularly susceptible to the illusion of control and to exaggerated optimism Let us grant, however, that most people do regard themselves as being happier and higher on their abilities, achievements, degree of control, and future prospects than they do most others, even most others from a similar socio-economic background Does it follow, as Taylor and Brown argue in (i) and (iii) above, that all of them, or even a majority, are laboring under an illusion? It does not. Depending on the figures, it could well be that only a small minority is laboring under an illusion.[27]

---

[25] For this study, Taylor and Brown cite J Freedman, *Happy People What Happiness Is Who Has It, and Why* (New York Harcourt Brace Jovanovich, 1978) (see IW 198)

[26] Taylor and Brown also state that most people believe that they can control an outcome even in situations of chance Colvin and Block question this claim by pointing out that in the studies they cite, "the difference between the depressed and non-depressed groups in regard to the illusion of personal control existed only at the outset of the experiment" (p 8) Colvin and Block surmise, correctly I believe, that the non-depressed subjects' illusion at the outset of the experiment "stems from the application by these generally effective subjects of their typical expectations to an unfamiliar and novel situation " When these subjects learn the nature of the situation, the difference between them and the depressed subjects disappears Perhaps most damning, Colvin and Block cite research by V A Benassi and H I M Mahler ("Contingency Judgments by Depressed College Students Sadder, But Not Always Wiser," *Journal of Personality and Social Psychology*, xxxv (1985) 917–27) that shows that "[s]light experimental variations can reverse the illusion of control effect such that depressed people overestimate their control of the situation and nondepressed people provide accurate estimates of control" (p 8) Last but not least, even if Taylor and Brown are right that nondepressed people tend to believe they can control situations where they have no control, it would not be germane to their thesis, since such an illusion would be extreme, not mild

Let us suppose that 60% think that they are happier, better, and have a brighter future than 60%. Only 20% of them *have* to be wrong, for it is certainly possible for 40% to outshine the other 60% in all these respects. So, only 20% of the population could be laboring under positive illusions. In the absence of a figure for what the respondents have in mind when they compare themselves with "most people," Taylor and Brown cannot conclude from the data that most people are systematically, even if mildly, deluded about themselves. The same considerations apply to argument (iv), according to which people's optimism about the future is not supported by base rate data: depending on the figures, it is possible that only a small minority is overly optimistic.

Everyday observation also suggests that most people are neither unrealistically positive about themselves across the board, nor overly optimistic about the future most of the time. Most of us are realistic in some domains of our lives, unrealistically optimistic or pessimistic in some others, and neither consistently realistic nor unrealistic in the rest. Indeed, Taylor and Brown themselves start to qualify importantly their claim that most people live under the influence of positive illusions when they say that "there are ways in which people exhibit self-corrective tendencies over time" because of "the need to monitor reality effectively" (PI 25). In the rest of the passage, however, they reiterate their view that most people are given to positive illusions most of the time, especially when "they are in an implemental mindset, attempting to put a decision into effect" (PI 25–26).

To return to their arguments: (ii) relies on the fact that most people rate themselves more positively on their attributes than do observers (IW 195–96). But why assume that the observers are more accurate than the subjects?[28] If most people tend to think more highly of themselves than of others, it could be not that they are inflating their own merits vis-à-vis the facts, but that they are deflating other people's

---

[27] Colvin and Block also point out that some of the respondents could, in fact, be happier and healthier than most people, but do not draw the conclusion that the vast majority of them could be (p. 14).

[28] Colvin and Block also question this assumption, noting that one of the empirical studies cited by Taylor and Brown for the discrepancy between self-ratings and observer ratings used undergraduate observers, and that "the observational basis for the ratings" was brief, leading to an extremely low agreement on the four negative attributes among the observers, and only a .49 on the seventeen desirable attributes (p. 8). Colvin and Block also reveal other problematic features of the study in question (P. M. Lewinsohn, W. Mischel, W. Chaplin, and R. Barton, "Social Competence and Depression: The Role of Illusory Self-Perceptions," *Journal of Abnormal Psychology*, LXXXIX (1980): 203–12), including the fact that although the subject groups consisted of (i) depressed subjects, (ii) "psychiatric control" subjects, that is, nondepressed individuals with other psychiatric

Suppose, however, that most people *are* systematically mildly deluded about themselves On what grounds do Taylor and Brown claim that they are mentally healthy or subjectively happy? One reason is simply that most people cannot be mentally unhealthy or unhappy (PI 22) [29] This conclusion, however, is too quick There are degrees of mental health and subjective happiness, just as there are degrees of physical health and of virtue People who are neither depressed nor psychotic (unhealthy) can still be dissatisfied or neurotic to varying degrees (less than fully healthy) The proliferation of therapists, counselors, gurus, and self-help books is evidence enough that dissatisfaction and neurosis are alive and well (a fact that also challenges the claim of widespread positive illusions) So whereas we can agree, for both empirical and conceptual reasons, that most people in a productive, vibrant society such as the U S cannot be insane, we have little reason to agree that most people are as healthy or subjectively happy (or virtuous) as they can realistically be

Let us grant, however, that most people are as healthy and subjectively happy as they can be On what do Taylor and Brown base their claim that this is so *because* of their illusions? They base it on the putative fact that the more realistic individuals tend to be depressed But as Colvin and Block have shown, research on the so-called "depressed realists" is contradictory or ambiguous Indeed, Taylor and Brown themselves acknowledge in "Positive Illusions and Well-Being Revisited" that several studies indicate that depressives are negatively biased towards themselves rather than realistic, and that there might *be* no "depressed realists" (PI 21)

---

problems, and (iii) normal controls, that is, nondepressed individuals without any psychiatric problems, the student observers were told that all the subjects were depressed (p 8) This negative bias led them to rate, on average, the "members of all three groups as below the midpoint on the dimension Lewinsohn's group labeled as "social competence," and, on average, below the ratings of "the members of all three groups" (p 8) But since depressed people have lower self-esteem, they rated themselves lower than the others, and were thus closer to the ratings of the observers The lower discrepancy, however, could simply be the result of two equally illusory negatively biased ratings, whereas the greater discrepancy in the case of the normals could be the result of greater objectivity on their part and negative illusions on the part of the undergraduate observers Colvin and Block also question the other studies and review articles used by Taylor and Brown to support their thesis

[29] Taylor and Brown mention only mental health here, but since they think that subjective happiness is a measure of health, it is safe to assume that they think their argument applies to subjective happiness as well

This renders Taylor and Brown's central thesis empty The thesis, it will be recalled, is comparative it states that mild positive illusions are *more conducive* to subjective happiness and mental health than realism So we need a group of unhappy or at least less happy realists to compare with the "normals," that is, the deluded Upbeat Taylor and Brown maintain that even if depressed people are not realistic, their original thesis remains intact[30] But if the Downbeat are unrealistic, *there is no example of depressed realists with whom to compare the Upbeat,* and hence no evidence for the thesis that positive illusions about oneself are *more* conducive to subjective happiness and health than realism The most Taylor and Brown can claim is that those who have *positive* illusions feel better about themselves (unsurprisingly), and are nicer and more creative, than those who have *negative* illusions about themselves But to say that mild *positive* illusions are better for us in these respects than mild *negative* illusions is hardly a point against realism it is a point against negative illusions vis-à-vis positive illusions (If you have to drink poison, better sweet poison than bitter)

For the sake of argument, however, let us assume that future research gives us evidence for the existence of depressed realists Taylor and Brown's thesis that mildly deluded people are subjectively happier and healthier than realistic people still faces problems

II 3 The thesis that positive illusions promote subjective happiness is based largely on self-reports But how reliable are these self-reports? Like most people, the subjects also likely believe that subjective happiness is, to some extent, an achievement, especially when it is understood (as it often is) not only as contentment but also as a sense of enjoyment and meaning But people who are given to illusions about their other achievements will also be given to illusions about their achievement of subjective happiness Indeed, we have some of the same reasons for thinking that the subjects inflate their subjective happiness as for thinking that they inflate their abilities and other achievements *they rate themselves as happier than most people* (IW 198)[31] Although Taylor and Brown faithfully report this finding, they fail to notice that it undermines their thesis that positive illusions promote subjective happiness rather than simply the *illusion* thereof.[32]

---

[30] At this point, however, Taylor and Brown reinterpret their original thesis as the relatively innocuous view that "Most healthy adults are positively biased in their self-perceptions" (PI 22) But given the view that there are degrees of health, this is hardly a challenge to the traditional view that we would be healthier without illusions

[31] As I argued in section II 2, this in itself need not show anything more than that a small minority is deluded about their achievements etc My point here is that, whatever the number of deluded people, we have the same grounds for being skeptical about their self-reported subjective happiness as about their other self-reported achievements

There is also a deeper, conceptual reason why people who harbor positive illusions about themselves must also harbor illusions about their level of subjective happiness. The reason is that the emotions and evaluations that express or constitute their illusions about their abilities, achievements, and future prospects—a sense of self-worth and personal efficacy, and a sense that the future is bright and open to their endeavors—together entail a sense of meaning and enjoyment of life. To the extent that one's life lacks meaning or enjoyment, one must either lack a sense of self-worth or personal efficacy, or lack hope for the future. It follows, then, that insofar as subjective happiness consists of these illusory emotions and evaluations, the connection between subjective happiness and illusions is a conceptual, and not a causal, one. In other words, people who are deluded about their abilities, achievements, degree of control, and future prospects *must* be deluded about their subjective happiness.[33] Taylor and Brown fail to see this because they fail to see that subjective happiness is not merely a sense of contentment.

In short, the research on which Taylor and Brown base their case does not show that positive illusions promote subjective happiness. To the extent that people's positive illusions are due to self-deception and subjective happiness is an achievement, there are no empirical grounds for thinking that people's positive illusions lead to greater subjective happiness rather than simply to an *illusion* of greater subjective happiness, and to the extent that subjective happiness consists of the positive emotions and evaluations that express or constitute these illusions, the connection between their subjective happiness and their illusions is a conceptual, and not a causal, one.

---

[32] Some readers have objected that an illusion of subjective happiness, even if based on self-deception, is no different from genuine subjective happiness. But the logical structure of self-deceptive happiness is no different from that of self-deceptive beliefs. If someone believes *p* self-deceptively, it follows that deep down he does not *really* believe *p*. Likewise, if someone deceives himself about (the extent of) his subjective happiness, it follows that deep down he is not *really* (that) subjectively happy. At best, he sort-of believes *p* and sort-of feels happy, while also believing not-*p* and feeling not-happy, respectively. Or perhaps his feelings and beliefs are unstable, "evanescent," subject to sudden shifts, as self-deceptive feelings and beliefs often are (Sartre, "Self-Deception," in Walter Kaufmann, ed., *Existentialism from Dostoevsky to Sartre* (New York: New American Library, 1975), pp. 302–03).

[33] See also Colvin and Block: "it is not surprising (and perhaps is even tautological or definitional) that individuals high on self-esteem, with a sense of control of their lives, and with optimism about the future are happier than individuals lacking these self-evaluations" (p. 16).

II 4 Taylor and Brown also hold that the Upbeat's positive illusions enable them to be more creative or productive, more able "to grow, develop, and self-actualize," more open to "new ideas and people," and more caring of other people than the Downbeat. But this thesis is subject to the same two counterarguments insofar as the evidence for their greater creativity and so on is based on the Upbeats' self-reports. Taylor and Brown do cite additional research to support their thesis, but this research provides only "suggestive evidence" that the ability to be creative, caring, open, and so on, are promoted by positive illusions about oneself (IW 200).

Let us suppose, however, that the credibility of the factual claim is not at issue, and the Upbeat really are all that Taylor and Brown claim them to be. The question now is why we should accept the causal claim that it is *because* of their positive illusions that the Upbeat have these features. Taylor and Brown's argument for this causal claim relies on the prior claim that realistic people are less likely to have these features, because they are more likely to be depressed and lack a sense of self-worth. It is certainly true that if we are down in the dumps about our worth or competence, it undercuts the motivation to be creative, to be open to growth and change, and to be caring of others.[34] In such a state, new ideas and people are likely to be threatening. But should we accept the claim that those who are down in the dumps about themselves are down because of their greater realism? So far I have proceeded on the assumption that further research will provide empirical evidence for this view. It is time now to question it.

II 5 The view in question is that the Downbeat are more realistic than the Upbeat, and depressed because of their realism. But if the Downbeat really are more realistic about their lives (and do not suffer from a naturally depressive temperament that prevents them from feeling happy, regardless of the facts), then it must be that the truth about their lives really is depressing. Perhaps they have been slothful or weak or pusillanimous, or perhaps they have only been incredibly unfortunate in their circumstances. In that case, however, their depression is due not to realism as such, but to the sad reality of their lives; after all, if this reality were positive, being realistic about their lives would be uplifting, not depressing. True, if they were extremely deluded about themselves, they could see their sloth, weakness, or pusillanimity as evidence of their good-natured, laid-back, live and let-live attitude towards life and, thus, they could feel good about themselves But extreme self-delusion, Taylor and Brown agree, is incompatible with the ability to grow and change and other criteria of mental health, criteria that their mildly self-deluded subjects apparently meet If, however, this reality is positive and they still come down hard on themselves, then it must be either that they have negative illusions about themselves, or that they see the facts for what they are but have unrealistic expectations of themselves

---

[34] This is the sort of psychological truism that leads Friedrich Nietzsche to declare in one of his famous—and famously one-sided—aphorisms that it is not virtue that produces happiness, but happiness that produces virtue. See Nietzsche, *Twilight of the Idols/The Anti-Christ* (1889/1895), R.J. Hollingdale, trans. (New York: Penguin, 1968), pp. 58–59.

The depression and low self-esteem of the Downbeat, then, can have only three explanations they have a naturally depressive temperament, they are realistic, but have nothing to be happy about, or, they have something to be happy about, but lack the realism to see it or to evaluate it by realistic standards Insofar as they lack this last, they lack the wisdom to accept either their own particular limitations or the limitations of human nature or both In none of these cases is it realism that is at fault, indeed, in the last case greater realism is essential for lifting their depression

The upshot is that, even if further research were to show a high correlation between realism, on the one hand, and depression, low self-esteem, and lack of creativity, and so on, on the other, Taylor and Brown would not be justified in drawing the conclusion that the depression and low self-esteem were a result of too much realism Without this premise, however, Taylor and Brown have no grounds for asserting that the greater subjective happiness, creativity, openness to new ideas and people, and so on, of the Upbeat are due to their positive illusions Indeed, not only do Taylor and Brown lack the empirical basis they need for their claim that positive illusions are more likely to promote happiness and mental health than realism, they have empirical grounds for believing the exact opposite For there is some evidence that some people are *both* more realistic *and* happier and healthier than Taylor and Brown's normals[35] Surprisingly, however, Taylor and Brown have nothing to say about them

In any case, we can see independently of empirical research why realistic self-perceptions are more conducive to mental health than illusory ones Growth and change are motivated by a perception of one's shortcomings, or by the aspiration to realize one's full potential by challenging oneself, and openness to new people and ideas is made more likely by a secure sense of self. To the extent that people are given to rosy illusions about their traits, abilities, or achievements, they lack the first motivation for growth and change, and to the extent that they have rosy illusions about their potential-fulfillment, they lack the second motivation for change. Moreover, to the extent that these illusions are due to self-deception rather than inculpable ignorance, self-deceived people's psychological investment in their illusions creates barriers against ideas and people who might puncture these illusions. Hence, whatever openness to growth and change people with positive illusions possess must be *in spite of* these illusions, not *because* of them.[36]

---

[35] Colvin and Block p 7, cite studies by W C Compton, "Are Positive Illusions Necessary for Self-Esteem A Research Note," *Personality and Individual Differences*, XIII (1992) 1343–44, and J Block and H Thomas, "Is Satisfaction with Self a Measure of Adjustment?," *Journal of Abnormal and Social Psychology*, LI (1955) 254–59

It is important to note that there are deep psychological and moral differences between those who are given to self-deception and those who are fundamentally reality-oriented but ignorant of themselves. It is a commonplace that knowledge of the external world is a life-long enterprise requiring experience, attention, and study. But knowledge of the self is no different in this respect, and so it is not surprising that most of us have large areas of ignorance about ourselves, and no one is completely self-knowledgeable. And what is true of most adults is even truer of college students. But people whose illusions are chiefly due to lack of experience rather than self-deception can still be fundamentally reality-oriented and, thus, open to new ideas and people, to growth and change. In such people, illusions are always in process of being whittled away. In lumping everyone with positive illusions together, Taylor and Brown fail to appreciate these differences.

Highly realistic people—people who tend to see themselves and others as they are, and who appraise themselves and others by realistic standards—are motivated to grow and change both by an awareness of their shortcomings, and by a desire to challenge themselves. Free of façade and defensiveness, they are unafraid of the truth and, thus, open to the challenge of new ideas and people. Such people are also highly likely to be subjectively fulfilled. If Taylor and Brown were right, however, it would follow that, other things being equal, highly realistic, healthy, and subjectively fulfilled people would tend to become healthier (and subjectively happier) if they became slightly deluded about themselves. It would also follow that healthy, subjectively fulfilled, but somewhat deluded people would tend to become less healthy (and less fulfilled) if they became more accurate in their perceptions as well as more realistic in their expectations and standards But this amounts to saying that people who have less reason to be open to new ideas and people, and who are thus less likely to grow and change, are more likely to be open to new ideas and people, and thus more likely to grow and change But this claim, of course, is incoherent Hence it is not only unsupported by any empirical evidence, it is immune to empirical evidence

---

[36] Colvin, Block, and Funder provide empirical support for the claim that people with positive illusions about themselves show "poor social skills and psychological maladjustment 5 years before and 5 years after the assessment of self-enhancement" (p. 1152).

II 6 One more claim needs to be addressed, namely, that unrealistic optimism among terminally ill patients leads to a greater (unrealistic) sense of control and a reduction in depression, pain, and disability (PI 24)

Taylor and Brown's discussion focuses on a study by Taylor and others of gay men at risk of AIDS [37] In this study, men who test positive for the HIV virus apparently show more optimism about not developing AIDS in their responses to the AIDS-specific optimism (ASO) questionnaire than men who test negative The illusory nature of their optimism is shown by their answers on the AIDS-specific optimism (ASO) questionnaire, in which some of them agree to such propositions as "I feel safe from AIDS because I've developed [or if exposed would develop] an immunity" to AIDS, and "I think my immune system would be (is) more capable of fighting the AIDS virus than that of other gay men" (*ibid*, p 469)

Overall, however, both the HIV-positive and the HIV-negative men disagree with the optimistic statements on the six-item ASO questionnaire, with means for the former at 1 94 and for the latter at 1 72 on a scale of 1 to 5 (*ibid*, p 463) [38] So it is misleading to claim that the HIV-positive men's responses show greater optimism rather than simply a little less pessimism Moreover, even the lesser pessimism is contradicted by the results of various other self-report measures, according to which, as the Taylor 1992 study reports, HIV-positive men "see themselves to be at greater risk of developing AIDS," have less of a sense of control over AIDS, and have more worries than the HIV-negative men (*ibid*, p 472) It cannot be true both that HIV-positive men are less pessimistic about not acquiring AIDS and that they "have higher levels of AIDS-related worries and concerns" than HIV-negative men, or both that they have "greater feelings of control" and that they "see themselves as having less control over AIDS" than HIV-negative men Perhaps the HIV-positive men's less pessimistic responses to the

---

[37] Shelley E Taylor, Margaret E Kemeny, Lisa G Aspinwall, Stephen G Schneider, Richard Rodriguez, and Mark Herbert, "Optimism, Coping, Psychological Distress, and High-Risk Sexual Behavior among Men at Risk for Acquired Immunodeficiency Syndrome (AIDS)," *Journal of Personality and Social Psychology*, LXIII, 3 (1992) 460–73

[38] See Colvin and Block, p 11

ASO questionnaire reflect their state of mind when they respond to statements that express optimism, and their more pessimistic responses to statements about AIDS-related worries reflect their state of mind when they respond to statements that express pessimism. But they cannot be both less and more pessimistic overall than the HIV-negative men.

But let us assume that the HIV-positive men are somewhat less pessimistic overall than the HIV-negative men, and ask why this might be so. Taylor and her colleagues offer the plausible explanation that it is because the HIV-positive men have decided to try and prolong their lives by avoiding risky behavior and engaging in various coping techniques (*op. cit.*, p. 472). But this, of course, is the reverse of the thesis Taylor's 1992 study (*op. cit.*, pp. 460, 470) and Taylor and Brown (PI 21, 24) set out to defend, namely, that the positive behavior and reduced distress are the *result* of their lesser pessimism. And although it is possible for the lesser pessimism about developing AIDS, on the one hand, and the efforts to prevent it, on the other, to become mutually reinforcing, each cannot be the initial cause of the other.

At any rate, those who falsely believe that they are immune to AIDS have no reason to do anything to protect themselves from it. So if they do try to protect themselves from it, it must be that they do not really, truly believe that they are immune—they are merely indulging in a bit of inert magical thinking when answering the ASO questionnaire, thinking that has no effect on their behavior.[39] The other clearly illusory belief, that they can eliminate the virus from their system if they take certain actions, *can* lead them to take those actions—but so can the true belief that they might be able to prevent, or at least delay, AIDS, if they take certain actions, a belief that they also hold.[40] In addition, it is well known that having a positive attitude has health-promoting benefits of its own. Consequently, if they adopt a positive attitude because they believe this, then their positive attitude is realistic. So while the false belief and an unrealistically positive attitude *can* produce the desired result, they are not *more conducive* to this result than the true belief and realistic optimism. Moreover, to the extent that people genuinely hold the false belief that they are immune to AIDS or that they can eliminate the virus from their system, they are likely to feel more unhappy when the disease catches up with them They are also at greater risk of leaving important projects incomplete or undone financial affairs neglected, frayed relationships unpatched, amends unmade, goodbyes unsaid, and so on And all this can only add to their unhappiness

---

[39] In situations of extreme stress, there is often a tendency to engage in magical thinking ("I'm probably immune to AIDS," "He's not really dead—I'll see him when I wake up in the morning") even while acting on realistic beliefs.

[40] As shown by the HIV-positive men's low score (1.95) on the Fatalistic Vulnerability scale (Taylor et al., p. 465). This score shows that most HIV-positive men strongly disagree that they will definitely develop AIDS, or that the spread of AIDS is inevitable (Taylor et al., p. 471). Their disagreement with these statements indicates that they believe (truly) that they can do something about their condition.

Taylor and her colleagues grant the possibility that the "optimists" will feel more unhappy if they develop AIDS But they also argue that, since "seropositive optimists as well as pessimists *readily acknowledge their risk and AIDS-related worries and concerns,* it seems unlikely that they will be greatly surprised if they subsequently develop AIDS" (PI 472, italics mine) Moreover, they continue, "inasmuch as optimism has been consistently associated with superior coping it may be that *the optimists will simply shift their expectations to accommodate the new reality* " (PI 472)

Given their view that the optimists are unrealistic, however, this argument borders on incoherence If the optimists are as ready to acknowledge the realities of their condition as the pessimists, and better able "to accommodate the new reality" than the pessimists, then they must be more realistic than the pessimists, even if they occasionally espouse some bizarre beliefs In any case, since very few HIV-positive men espouse bizarre beliefs about their condition, most of those who cope well must do so with the help of realistic optimism Once again, however, realistic optimists are never even mentioned by Taylor and Brown or Taylor's 1992 study Before concluding, then, it would help to have an example that shows vividly how realistic optimism can help a person not only cope with a terrible disease, but even use it to enrich her life

II 7 In a radio essay called "I Always Have a Choice," Catherine Royce, a dancer for thirty years, relates that in 2003 she discovered that she had ALS (Lou Gehrig's disease) [41] Facing the facts unflinchingly, she states that over time ALS will "destroy every significant muscle in my body Ultimately, I will be unable to move, to speak and, finally, to breathe " But "I believe I always have a choice No matter what I'm doing No matter what is happening to me I always have a choice " When she loses the power to type, she has the choice to give up writing—or to learn to use voice recognition software She does the latter and ends up writing "more now than ever " She also has the choice to live—or to die "I can choose to see ALS as a death sentence, or I can choose to see it as an invitation—an opportunity to learn who I truly am " She does the latter, discovering "an ability to recognize, give and receive caring in a way far deeper than anything in my life before " Yet she does not pretend that ALS is a blessing in disguise, and she does not have any illusions of control over it She is optimistic, though fully aware of the terrible course of her disease, and she knows that she always has a choice, though fully aware that her choices are restricted Her life is both objectively worthwhile and, so far, subjectively fulfilling

---

[41] National Public Radio, December 4, 2006, http //www npr org/templates/story/ story php?storyId=6560320&sc=emaf

III CONCLUSION

I started by outlining the widespread idea that (*eudaimonic*) happiness is subjective fulfillment in an objectively worthwhile life, and that realism is an important means to subjective fulfillment and objective worth, and conceptually necessary for objective worth Realism requires being reality-oriented, in touch with the important facts of one's life and human life in general, and disposed to evaluate oneself and others by realistic standards The claim that realism is an important means to subjective happiness faces challenges both from common sense and from social psychology I have defended it against both challenges Common sense is right in holding that some truths can be disastrous for us, robbing us of subjective well-being as well as the ability to respond in worthwhile ways, and that being reality-oriented is more likely to lead us to such truths In such cases, some subjective happiness based on illusion is better than utter misery based on truth Being self-deceived, however, is far more likely to lead to disaster My theory also recognizes that the two components of *eudaimonic* happiness, subjective fulfillment and objective worth, can conflict, making such happiness difficult to achieve

I have given several reasons for rejecting the Taylor-Brown thesis that mild positive illusions are more likely to promote subjective happiness than realism (i) Contrary to their claim, empirical research gives little reason to believe that most people are given to mild positive illusions about themselves Rather, experience suggests that most people harbor (mostly) positive illusions in some areas of their lives, (mostly) negative illusions in some others, and (mostly) realism in yet others (ii) Even if most people are mildly deluded, to the extent that their delusions are due to self-deception rather than simple ignorance, we cannot, for both empirical and conceptual reasons, use their self-reports to claim a causal connection between their illusions and their subjective happiness These reasons apply as well to the Upbeat's self-reports of greater mental health The remaining empirical evidence for a causal connection between positive illusions and health is only "suggestive " (iii) There is no good evidence for depressed realists, so the thesis that positive illusions are *more* conducive to subjective happiness and health than realism is vacuous (iv) Since realism, properly understood, requires seeing ourselves accurately and appraising ourselves by realistic standards, even if depressed realists existed, their depression could not be due to their realism (v) Empirical research gives evidence of realists who are happier and healthier than people with positive illusions (vi) The claim that positive illusions about oneself promote creativity, growth, and so on, even if one is self-deceptively invested in those illusions, defies explanation, whereas the opposite claim is easily explained (vii) Neither empirical research nor philosophical argument supports the idea that illusory optimism about AIDS helps people cope better with their condition than realistic optimism (viii) The example of Catherine Royce shows how realistic optimism can help a person not only cope with her condition but even enrich her life Royce is an exemplar of the courageous, life-affirming, well-functioning individual that philosophers and psychologists from Aristotle to Maslow and Rogers have held up as models for us to aspire to It is hard to see how any illusion about her condition or her future prospects could possibly contribute to her health or happiness

NEERA K BADHWAR

University of Oklahoma

NEW BOOKS

GARVER, EUGENE *Confronting Aristotle's Ethics* Chicago University Press, 2006 ix, 328 p Cloth $49 00

GASKIN, RICHARD *Experience and the World's Own Language A Critique of John McDowell's Empiricism* New York Oxford, 2006 xi, 251 p Cloth $65 00

GASTON, SEAN *The Impossible Mourning of Jacques Derrida* New York Continuum, 2006 viii, 152 p Cloth $24 95

GAUKER, CHRISTOPHER *Conditionals in Context* Cambridge MIT, 2005 xi, 333 p Paper $35 00

GAUKROGER, STEPHEN *The Emergence of a Scientific Culture Science and the Shaping of Modernity 1210–1685* New York Oxford, 2006 ix, 563 p Cloth $65 00

GAUT, BERYS *Art, Emotion and Ethics* New York Oxford, 2007 269 p Cloth $70 00

GAUTHIER, DAVID *Rousseau The Sentiments of Existence* New York Cambridge, 2006 xiv, 196 p Cloth $60 00, paper $22 99

GEORGALIS, NICHOLAS *The Primacy of the Subjective Foundation for a Unified Theory of Mind and Language* Cambridge MIT, 2005 vii, 352 p Cloth $36 00

GEUSS, RAYMOND *Outside Ethics* Princeton University Press, 2006 viii, 257 p Cloth $59 50, paper $24 95

GIBBS, RAYMOND W , JR *Embodiment and Cognitive Science* New York Cambridge, 2006 336 p Cloth $70 00, paper $25 99

GIERE, RONALD N *Scientific Perspectivism* Chicago University Press, 2006 151 p Cloth $30 00

GIESECKE, ANNETTE LUCIA *The Epic City Urbanism, Utopia, and the Garden in Ancient Greece and Rome* Wasington DC Center for Hellenic Studies, 2007 xiv, 204 p Paper $18 95

GILBERT, MARGARET *A Theory of Political Obligation Membership, Commitment, and the Bonds of Society* New York Oxford, 2006 x, 332 p Cloth $55 00

GILL, MICHAEL B *The British Moralists on Human Nature and the Birth of Secular Ethics* New York Cambridge, 2007 Cloth $85 00

GIAQUINTO, MARCUS *Visual Thinking in Mathematics An Epistemological Study* New York Oxford, 2007 vii, 287 p Cloth $72 00

GIORDANO, SIMONA *Understanding Eating Disorders Conceptual and Ethical Issues in the Treatment of Anorexia and Bulimia Nervosa* New York Oxford, 2005 xiii, 297 p Cloth $85 00

GLAD, JOHN *Future Human Evolution Eugenics in the Twenty-First Century* Schuylkill Haven, PA Hermitage, 2006 136 p Cloth NP

GLENDINNING, SIMON *The Idea of Continental Philosophy* Edinburgh, UK. University Press, 2007 144 p Cloth $75 00, paper $22 00

GLOVER, JONATHAN *Choosing Children The Ethical Dilemmas of Genetic Intervention* New York Oxford, 2006 120 p Cloth $15 95

GOLDMAN, ALVIN I *Simulating Minds The Philosophy, Psychology, and Neuroscience of Mindreading* New York Oxford, 2006 x, 364 p Cloth $35 00

GOOD, JUSTIN *Wittgenstein and the Theory of Perception* New York Continuum, 2006 vi, 186 p Cloth $120 00

GOODPASTER, KENNETH E *Conscience and Corporate Culture* Malden, MA Blackwell, 2007 xxvi, 306 p Paper $24 95

THE JOURNAL OF PHILOSOPHY

Special Issues Available

Parts and Wholes — December 2006
On Bernard Williams — June 2003
Causation — April 2000
Jurgen Habermas and John Rawls — March 1995

The Dewey Lectures.
Michael Dummett — *Truth and the Past*, 2002
Hilary Putnam — *Sense, Nonsense, and the Senses*
An Inquiry into the Powers of the Human Mind, 1994
Donald Davidson — *The Structure and Nature of Truth*, 1990
Amartya Sen — *Well-being, Agency, and Freedom,* 1985
John Rawls — *Kantian Constructivism in Moral Theory*, 1980
Ernest Nagel — *Teleology Revisited,* 1977
W V Quine — *Ontological Relativity*, 1968

The Woodbridge Lectures
John McDowell - *Having the World in View*
Sellars, Kant, and Intentionality, 1997

Subscriptions

Individuals $45 00
Students $20 00
Libraries and Institutions $100 00
Print and Online for Libraries and Institutions $180 00
Postage outside the United States $15 00

Special Offer

One complimentary issue of any of the above
issues with each new subscription
All issues are $5 00 plus $3 00 for shipping and handling
Send check or money order to
The Journal of Philosophy
Columbia University
1150 Amsterdam Avenue, Mail Code 4972
New York, NY 10027

www journalofphilosophy org

Young Epistemologist Prize

The **Rutgers Epistemology Conference** announces the sixth bi-annual prize for the best essay submitted to the Conference by a person with a Ph D obtained by the time of submission but not earlier than ten (10) years prior to the date of the conference Thus, for the Conference on May 8th/9th, 2009, the Ph D must not have been awarded prior to May, 1999 The prize winning essay will be presented at the bi-annual conference and will be published (along with the invited papers) in Philosophy and Phenomenological Research The winner of the prize will receive an award of $1,000 plus all travel and lodging expenses connected with the Conference

Details

The essay may be in any area of epistemology The essay ***must*** be limited to 5,000 words (not including footnotes and bibliography) and submitted on a PC-formatted disk along with three (3) printed copies The essays will be judged blindly and, thus, the three printed copies (but not the disk copy) must be prepared in such a way that the author's identity is completely masked ***Only*** papers submitted in that fashion will be considered Papers ***must*** be postmarked by November 15, 2008 The winner of the prize will be announced by February 15, 2009 Please include your **email address** in the cover letter accompanying your paper By submitting the essay, the author agrees to have it posted on the Rutgers Epistemology Conference Webpage and published in Philosophy and Phenomenological Research

Please send all papers (along with the disk and cover letter) to

The Young Epistemologist Prize
Philosophy Department
Rutgers University
26 Nichol Ave
New Brunswick, NJ 08901-2882

For more information about the Rutgers Epistemology Conference and the Young Epistemologist Prize, please go to the Conference Website

http://epistemology.rutgers.edu

New from OXFORD

Drugs and Justice

Seeking a Consistent, Coherent, Comprehensive View

MARGARET P BATTIN, ERIK LUNA, ARTHUR G LIPMAN, PAUL M GAHLINGER, DOUGLAS E ROLLINS, JEANETTE C ROBERTS, and TROY L BOOHER

This compact and innovative book tackles one of the central issues in drug policy the lack of a coherent conceptual structure for our thinking about drugs Battin and her contributors lay a foundation for a wiser drug policy by promoting consistency and coherency in the discussion of drug issues and by encouraging a unique dialogue across disciplines

2007 312 pp, 10 line illus cloth $74 00 paper $21 95

David Hume

A Dissertation on the Passions, The Natural History of Religion

Edited by TOM BEAUCHAMP

Tom Beauchamp presents the definitive scholarly edition of two famous works by David Hume, both originally published in 1757 Authoritative critical texts are accompanied by a full array of editorial matter

(The Clarendon Edition of the Works of David Hume)

2008 456 pp $130 00

New in Paperback

The Oxford Handbook of Contemporary Philosophy

FRANK JACKSON and MICHAEL SMITH

The Oxford Handbook of Contemporary Philosophy is the definitive guide to what's going on in this lively and fascinating subject The coverage is broad, with sections devoted to moral philosophy, social and political philosophy, philosophy of mind and action, philosophy of language, metaphysics, epistemology, and philosophy of the sciences

(Oxford Handbooks in Philosophy)

2008 920 pp paper $49 50 cloth $125 00

The Emotional Construction of Morals

JESSE PRINZ

Jesse Prinz presents a bravura argument for highly controversial claims about morality, which go to the heart of our understanding of ourselves

2008 320 pp $49 95

Reification

A New Look At An Old Idea

AXEL HONNETH and MARTING JAY

Can the concept of reification inspire 21st-century political theory? Axel Honneth, answers by drawing on his theory of recognition and then responds to three eminent critics Judith Butler, Raymond Geuss, and Jonathan Lear

(The Berkeley Tanner Lectures)

2008 184 pp $29 95

Law, Reason, and the Cosmic City

Political Philosophy in the Early Stoa

KATJA MARIA VOGT

This book argues that political philosophy is central to early Stoic philosophy, and is deeply tied to the Stoics' conceptions of reason and wisdom Broad in scope, it explores the Stoics' idea of the cosmic city, their notion of citizen-gods, as well as their account of the law

2008 256 pp $55 00

Trials of Reason

Plato and the Crafting of Philosophy

DAVID WOLFSDORF

The study focuses on a set of fourteen so-called early dialogues, beginning with a methodological framework that explains how to integrate the argumentation and the drama in these texts *Trials of Reason* argues that Plato's early dialogues are as much works of meta-philosophy as philosophy itself

2008 296 pp, 4 line illus $74 00

Prices are subject to change and apply only in the US To order, or for more information, please call 1-800-451-7556 In Canada, call 1-800-387-8020 Visit our website at www oup com/us

OXFORD
UNIVERSITY PRESS

New Titles From **Prometheus Books**

NIETZSCHE AND THE "ENGLISH"

The Influence of British and American Thinking on His Philosophy

Thomas H. Brobjer

474 pp / PB / $39.95 / ISBN 978-1-59102-547-4

REASON AND EMANCIPATION

Essays on the Philosophy of Kai Nielsen

Edited by Michel Seymour and Matthias Fritsch

431 pp / HC / $64 / ISBN 978-1-59102-418-7

C. S. LEWIS AND THE SEARCH FOR RATIONAL RELIGION

Revised and Updated

John Beversluis

363 pp / PB / $20 / ISBN 978-1-59102-531-3

THE SECULAR CONSCIENCE

Why Belief Belongs in Public Life

Austin Dacey

269 pp / HC / $24.95 / ISBN 978-1-59102-604-4

THE KALAM COSMOLOGICAL ARGUMENT FOR GOD

Mark R. Nowacki

315 pp / HC / $59 / ISBN 978-1-59102-473-6

NEW INTERPRETATIONS OF BERKELEY'S THOUGHT

Edited by Stephen H. Daniel

319 pp / HC / $80 / ISBN 978-1-59102-557-3

PHILOSOPHY OF BIOLOGY

Second Edition

Edited by Michael Ruse

473 pp / PB / $24 / ISBN 978-1-59102-527-6

ANIMALS, EMOTION, AND MORALITY

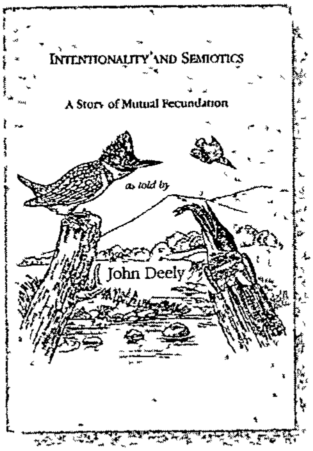
Marking the Boundary

B. A. Dixon

240 pp / PB / $21.95 / ISBN 978-1-59102-629-7

Prometheus Books
Toll free 800-421-0351 / Fax 716-691-0137
www.prometheusbooks.com

Philosophy from the University of Scranton Press



Intentionality and Semiotics

A Story of Mutual Fecundation

John N. Deely

John Deely offers a controversial solution to the problem of subjectivity in inquiry. He creates an interface between semiotics and the concept of intentionality, as it appears in Aquinas's work, to demonstrate that every sign is irrevocably linked to human understanding. In the process, Deely builds a bridge between classical thinkers such as Aristotle and modernists such as Heidegger and Wittgenstein.

Cloth $25.00

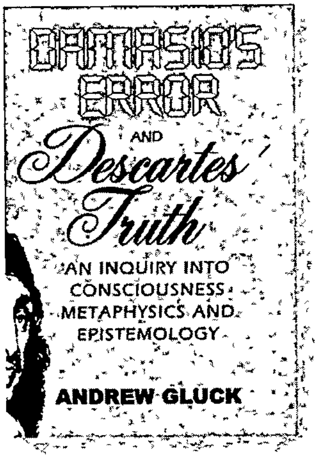
Damasio's Error and Descartes' Truth

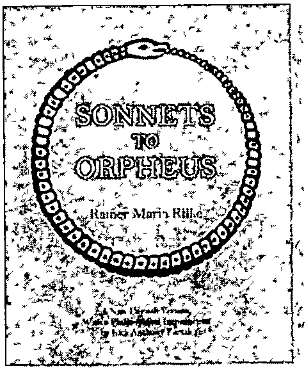
An Inquiry into Consciousness, Metaphysics, and Epistemology

Andrew Gluck

The question of the relationship between mind and body as posed by Descartes, Spinoza, and others remains a fundamental debate for philosophers. In *Damasio's Error and Descartes' Truth*, Andrew Gluck constructs a pluralistic response to the work of neurologist Antonio Damasio.

Paper $15.00


Damasio's Error and Descartes' Truth
An Inquiry into Consciousness, Metaphysics and Epistemology
Andrew Gluck



Sonnets to Orpheus

Rainer Maria Rilke

Translated and with an Introduction by
Rick Anthony Furtak

Rainer Maria Rilke's *Sonnets to Orpheus*, first published in German in 1923, has been hailed by critics as some of the finest lyrical verse of the twentieth century. Rick Anthony Furtak's translation of this legendary collection is the first edition for an English-speaking audience to situate the poems in a philosophical context, lending unexpected depth and a fresh perspective to Rilke's magnum opus.

Cloth $15.00

Distributed by **The University of Chicago Press** • *www.press.uchicago.edu*

AMBRIDGE

New and Noteworthy Titles

War
Essays in
Political Philosophy
Edited by Larry May
$85 00 Hb 978-0 521 87637 7
$28 99 Pb 978 0-521-70004-7
344 pp

Aggression and Crimes Against Peace
Larry May
$85 00 Hb 978-0-521-89431-9
$27 99 Pb 978-0-521-71915-5
300 pp

An Introduction to the Philosophy of Religion
Michael J Murray and Michael C. Rea
Cambridge Introductions to Philosophy
$90 00 Hb 978 0-521-85369-9
$29 99 Pb 978-0-521-61955 4
306 pp

God and Phenomenal Consciousness
A Novel Approach to
Knowledge Arguments
Yujin Nagasawa
$85 00 Hb 978-0-521-87966 8
176 pp

Mathematical Thought and Its Objects
Charles Parsons
$95 00 Hb 978 0-521 45279-3
398 pp

Toward a Theory of Human Rights
Religion, Law, Courts
Michael J Perry
$27 99 Pb 978 0-521 68422 4
240 pp

Choosing to Die
Elective Death and
Multiculturalism
C G Prado
$85 00 Hb 978-0-521-87484-7
$23 99 Pb 978 0 521-69758-3
224 pp

An Introduction to Non-Classical Logic 2nd Edition
From If to Is
Graham Priest
Cambridge Introductions to Philosophy
$99 00 Hb 978 0 521-85433-7
$34 99 Pb 978-0-521-67026 5
648 pp

Now in Paperback!
Darwinism and its Discontents
Michael Ruse
$19 99 Pb 978 0-521-72824-9
326 pp

Hobbes and Republican Liberty
Quentin Skinner
$70 00 Hb 978 0-521-88676-5
$22 99 Pb 978-0 521-71416 7
270 pp

I Was Wrong
The Meanings of Apologies
Nick Smith
$99 00 Hb 978 0-521-86552 4
$24 99 Pb 978 0-521-68423-1
304 pp

Evidence and Evolution
The Logic Behind the Science
Elliott Sober
$90 00 Hb 978 0-521-87188 4
$29 99 Pb 978-0-521-69274-8
412 pp

Conflict of Interest and Public Life
Cross-National Perspectives
Christine Trost and Alison L Gash
$80 00 Hb 978-0 521-88142-5
276 pp

Prices subject to change

www cambridge org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

New and Noteworthy Titles

Leibniz
An Intellectual Biography
Maria Rosa Antognazza
$39.95 Hb 978 0-521-80619 0
550 pp

God, Chance and Purpose
Can God Have It Both Ways?
David J. Bartholomew
$80.00 Hb 978-0-521-88085 5
$19.99 Pb 978 0-521-70708-4
272 pp

The Problem of Punishment
David Boonin
$85.00 Hb 978-0 521-88316 0
$27.99 Pb 978 0-521-70961-3
312 pp

Alfred Tarski
Life and Logic
Anita Burdman Feferman and Solomon Feferman
$24.99 Pb 978 0 521-71401 3
432 pp

Justice, Political Liberalism, and Utilitarianism
Themes from Harsanyi and Rawls
Edited by Marc Fleurbaey, Maurice Salles, and John A. Weymark
$85.00 Hb 978 0-521-64093-0
472 pp

What is Analytic Philosophy?
Hans-Johann Glock
$90.00 Hb 978-0-521-87267 6
$24.99 Pb 978-0-521 69426-1
304 pp

Anti-Individualism
Mind and Language, Knowledge and Justification
Sanford C. Goldberg
Cambridge Studies in Philosophy
$90.00 Hb 978-0-521-88048 0
280 pp

The Fall of Man and the Foundations of Science
Peter Harrison
$95.00 Hb 978 0-521 87559 2
312 pp

Interpreting Spinoza
Critical Essays
Edited by Charlie Huenemann
$90.00 Hb 978 0-521-87183-9
206 pp

Ethics and the Environment
An Introduction
Dale Jamieson
Cambridge Applied Ethics
$90.00 Hb 978-0 521-86421-3
$29.99 Pb 978-0-521-68284 8
234 pp

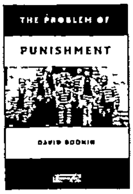
Ethics and Criminal Justice
An Introduction
John Kleinig
Cambridge Applied Ethics
$90.00 Hb 978 0-521-86420 6
$29.99 Pb 978-0-521-68283-1
294 pp

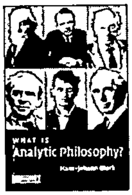
An Introduction to Buddhist Philosophy
Stephen J. Laumakis
Cambridge Introductions to Philosophy
$90.00 Hb 978 0 521-85413 9
$32.99 Pb 978-0 521-67008-1
300 pp

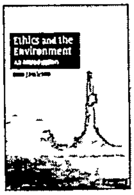
Science, Politics, and Evolution
Elisabeth A. Lloyd
Cambridge Studies in Philosophy and Biology
$85.00 Hb 978-0-521 86570-8
320 pp

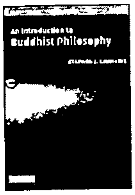
Anthropology, History, and Education
Edited and translated by Robert B. Louden and Gunter Zoller
The Cambridge Edition of the Works of Immanuel Kant in Translation
$150.00 Hb 978-0-521-45250-2
574 pp

Prices subject to change



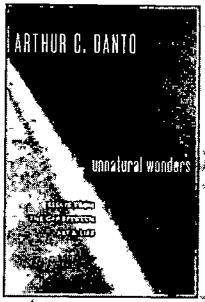


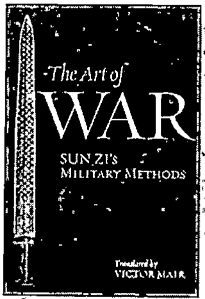


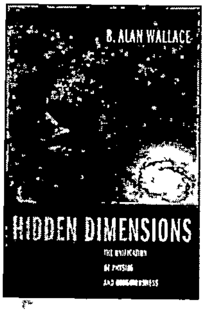


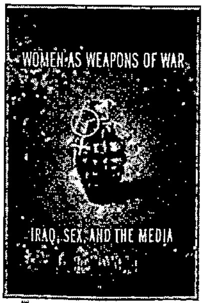
www.cambridge.org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS









COLUMBIA

Read book excerpts at www.columbia.edu/cu/cup

Unnatural Wonders
Essays from the Gap Between Art and Life
Arthur C. Danto
"Among the most sensible, intelligent, logical, and accessible art criticism of the last five years."
— Kirkus
paper $22.50

The Art of War
Sun Zi's Military Methods
Sun Zi (Sun Tzu)
Translated by Victor Mair
"[A] brilliant translation especially commendable for its clarity and elegance."
— Nicola Di Cosmo, *Princeton University*
cloth $19.95
Translations from the Asian Classics

The Politics of Our Selves
Power, Autonomy, and Gender in Contemporary Critical Theory
Amy Allen
cloth - $34.50
New Directions in Critical Theory

Crossing Horizons
World, Self, and Language in Indian and Western Thought
Shlomo Biderman
Translated by Ornan Rotem
cloth - $40.00

Columbia Companion to Twentieth-Century Philosophies
Edited by Constantin V. Boundas
cloth $75.00

Hidden Dimensions
The Unification of Physics and Consciousness
B. Alan Wallace
"Important and provocative."
— Arthur Zajonc, author of *The New Physics and Cosmology*
cloth $24.50
Columbia Series in Science and Religion

Women as Weapons of War
Iraq, Sex, and the Media
Kelly Oliver
"Immensely pleasurable and intellectually satisfying."
— Cynthia Willett, *Emory University*
cloth $26.50

Derrida, Deleuze, Psychoanalysis
Edited by Gabriele Schwab
paper - $24.50 / cloth $74.50

The Analects of Confucius
Burton Watson
cloth $19.95
Translations from the Asian Classics

Consciousness and Mental Life
Daniel N. Robinson
cloth $29.50

This Is Not Sufficient
An Essay on Animality and Human Nature in Derrida
Leonard Lawlor
cloth - $29.50

10.9.08
UNIVERSITY OF CALCUTTA

THE JOURNAL OF PHILOSOPHY

VOLUME CV, NUMBER 3
MARCH 2008

page
109 *Nobody Needs a Theory of Art* Dominic McIver Lopes
128 *What Colors Could Not Be: An Argument for Color Primitivism* Joshua Gert
COMMENTS AND CRITICISM
156 *Risky Business: Evolutionary Theory and Human Attitudes toward Risk—A Reply to Okasha* Armin W. Schulz
166 NEW BOOKS

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY
FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Michael J Mooney, *Treasurer,* Liaquat Ahamed, *Other Trustees* Lee Bollinger, Leigh S Cauman, Barbara Gimbel, Kent Greenawalt, Lynn Nesbit, Daniel Shapiro

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721
You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY
2008
SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices
POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2008 by the Journal of Philosophy, Inc ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY

VOLUME CV, NO. 3, MARCH 2008

CENTRAL LIBRARY
UNIVERSITY OF CALCUTTA
ALIPORE CAMPUS

NOBODY NEEDS A THEORY OF ART*

> It would never have occurred to the painters of Lascaux that they were producing *art* on those walls Not unless there were neolithic aestheticians
>
> —Arthur Danto

What is art? More precisely, what makes any item a work of art? The question is "probably the most venerable in aesthetics," according to Jerrold Levinson,[1] and Monroe Beardsley adds that it is "the most vexing "[2] In fact, the following answer to the question should be obvious

(R) item *x* is a work of art if and only if *x* is a work in activity *P* and *P* is one of the arts

The arts, of course, include music, dance, theater, literature, film, painting, architecture, and the like Yet (R) has seemed so far from obvious that, until now, nobody has given it a moment's thought The trouble is not that anyone might seriously deny its truth, but rather that they will find it uninformative After all, the vexing question is pressed upon us by radical changes in art of the avant-garde, and (R) offers no resources to address these changes With that in mind, here is the case for (R) The challenges posed by the avant-garde are real and they need to be addressed, but the vexing question is the wrong question to address them It does not follow that the question has no good answer On the contrary, (R) is all the answer we need, if

---

*Thanks to audiences at Lewis and Clark College, St Andrews University, the University of Kent, the University of Leeds, and the University of Lethbridge for helpful suggestions on earlier versions of this paper

[1] Levinson, "Defining Art Historically," *British Journal of Aesthetics*, xix (1979) 232–50, here p 232

[2] Beardsley, "Redefining Art," in *The Aesthetic Point of View* (Ithaca Cornell, 1982), pp 298–316, here p 298

0022-362X/08/0503/109–27

we do not need an answer that addresses the challenges posed by the avant-garde. Moreover, (R) points to questions that we do need answered. So, not only is it true but, in addition, (R) is as informative as we need.[3]

I. FROM KRISTELLER TO TESTADURA

A theory of art is an answer to the question of what makes an item a work of art. (R) is one such theory among many.[4] The case against (R) is that it fails to address the challenges posed by the avant-garde. The case for it is that theories of art are not needed to address the challenges of the avant-garde. However, the case for (R) runs immediately into a road block. The avant-garde is not much older than a century, but the vexed question is venerable, and so is the search for a theory of art. It therefore seems that we expect a theory of art to do more than cope with developments in the last century. Clearing this road block means doing a little history, and the history supplies materials useful in making the case for (R).

As far back as 1914, Clive Bell identified the "central problem of aesthetics" with the search for "the quality that distinguishes works of art from all other classes of objects."[5] Morris Weitz began his famous attack on Bell's project by admitting that "theory has been central in aesthetics and is still the preoccupation of the philosophy of art. Its main avowed concern remains the determination of the nature of art which can be formulated into a definition of it" (*op. cit.*, p. 27). Responses to Weitz play along, conceding that "underlying every traditional aesthetic theory is the essentialist presumption that that the expression 'work of art' applies to the entities that it does in virtue of some shared essential property or properties."[6]

---

[3] This paper deflects the vexed question by defending (R) while taking seriously the challenges of the avant-garde. Others dismiss the vexed question. For example, Kendall Walton, "Review of *Art and the Aesthetic*, by George Dickie," *Philosophical Review*, LXXXVI (1977): 97–101; Richard Wollheim, *Painting as an Art* (New York: Thames and Hudson, 1987), and Peter Kivy, *Philosophies of Arts* (New York: Cambridge, 1997).

[4] Notably Morris Weitz, "The Role of Theory in Aesthetics," *Journal of Aesthetics and Art Criticism*, XV (1956): 27–35; Monroe Beardsley, "An Aesthetic Definition of Art," in Hugh Curtler, ed., *What Is Art?* (New Haven: Yale, 1983), pp. 15–29; Arthur Danto, *Transfiguration of the Commonplace* (Cambridge: Harvard, 1981); George Dickie, *The Art Circle* (New York: Haven, 1984); Levinson, "Defining Art Historically"; Stephen Davies, *Definitions of Art* (Ithaca: Cornell, 1991); Berys Gaut, "'Art' as a Cluster Concept," in Noël Carroll, ed., *Theories of Art Today* (Madison: Wisconsin UP, 2000), pp. 25–44; and Gary Iseminger, *The Aesthetic Function of Art* (Ithaca: Cornell, 2004).

[5] Bell, *Art* (London: Chatto and Windus, 1914), p. 3.

[6] Robert Matthews, "Traditional Aesthetics Defended," *Journal of Aesthetics and Art Criticism*, XXXVIII (1979): 39–50, p. 39.

A different story is told by Paul Kristeller in his classic study, "The Modern System of the Arts "[7] According to Kristeller, European texts deploy no concept of art until the eighteenth century The Greek *techne* and Latin *ars*, for instance, refer to all kinds of crafts and sciences The eighteenth-century innovation is to group some activities apart from others For the first time in history, painting, sculpture, architecture, music, and poetry are grouped (sometimes with decorative art, landscape architecture, dance, theater, or prose literature) and also distinguished from the liberal arts, the practical arts, and the sciences Fishing, anatomical dissection, poetry, philosophy, football, travel, eating, architecture, fashion, trading, astronomy, dance, painting, plumbing which of these belong together? A grouping is salient for you and me that was not salient for contemporaries of Aristotle and Abelard Kristeller writes that "the various arts are certainly as old as human civilization, but the manner in which we are accustomed to group them and to assign them a place in our scheme of life and of culture is comparatively recent" (*ibid* , p 45)

Kristeller identifies some of the factors that, over the centuries, drove the innovation While subsequent scholarship challenges some of the details, Kristeller's account illustrates the kinds of forces that were in play The humanists gave poetry top spot over grammar and rhetoric in their new curriculum Painting, sculpture, and architecture gradually gained prestige from the fourteenth century onwards A new body of literature in the sixteenth century compared poetry and painting The seventeenth century saw the founding of the French academies, which sponsored the first treatises on painting, sculpture, and architecture to stand alongside texts on poetics Contributors to the quarrel between the ancients and the moderns distinguished activities where success depends on accumulated knowledge (science, where the moderns win) from activities where success depends on individual talent (the arts, where the ancients win) Thus the new concept of science helped spur the new concept of art Finally, the newly-drawn distinction between beauty and moral goodness suggested a domain for each By the early eighteenth century the "problem of the arts" was a topic of learned discussion in Paris and by mid-century a consensus settled on the modern system of the arts [8]

---

[7] Kristeller, "The Modern System of the Arts," *Journal of the History of Ideas*, xii (1951) 496–527, and xiii (1952) 17–46 See also Larry Shiner, *The Invention of Art A Cultural History* (Chicago University Press, 2001)

[8] Especially Charles Batteux, *Les Beaux arts réduits a un même principe* (Paris, 1746), and Denis Diderot, ed , *Encyclopédie ou Dictionnaire Raisonné des sciences, des arts, et des métiers* (Paris, 1751)

Several lessons are typically drawn from Kristeller's essay Some note that the concept of art is historically conditioned it is acquired at a certain time in a certain place and it spreads from there However, this is neither surprising nor interesting, for many concepts are historically conditioned in this sense Others go further, concluding that there is no art before the eighteenth century [9] This implies that there is no art at a time unless people have a concept of art at that time—a controversial claim [10] (Kristeller holds that art is ancient even if the concept of art is modern) A third lesson is that the concept of art begins as a theoretical concept Like the concepts of polymers and logical completeness, it is introduced by means of a theory, indeed a definition The concept of art and theories of art are equally venerable twins

Easily missed is a fourth lesson, which qualifies the third lesson True, the concept of art is from the get-go a theoretical concept devised by intellectuals Kristeller is clear about their goals They aimed to establish the unity of the arts—to make salient a scheme grouping together painting, sculpture, architecture, music, and poetry (and some others) They also sought to identify what sets these arts apart from other activities, especially the sciences and practical arts [11] They hoped to understand the distinctive features of each of the arts [12] Finally, some expected to ground rankings of the arts This is not at all the project of Bell and his successors Recent theories of art state what makes any given item a work of art The early moderns wanted to know what makes any given activity one of the arts They sought a theory of the arts

To see the difference, consider the form of each A theory of art completes the schema,

item *x* is a work of art if and only if

A theory of the arts completes the schema,

activity *P* is an art form if and only if [13]

---

[9] For example, Stephen Davies, "Definitions of Art," in Berys Gaut and Dominic McIver Lopes, eds , *Routledge Companion to Aesthetics* (New York Routledge, 2005, 2nd ed), pp 227–39, here p 234

[10] Lopes "Art without 'Art'," *British Journal of Aesthetics*, XLVII (2007) 1–15

[11] In *The Invention of Art*, Shiner stresses the distinction between the arts and the practical sciences, or crafts This emphasis might project a later concern onto the earlier period, when it was just as important to distinguish the arts from the sciences

[12] For example, G E Lessing, *Laocoon*, E A McCormick, trans (Indianapolis Bobbs-Merrill, 1962 [1766])

[13] Theories are definitions if they give individually necessary and jointly sufficient conditions They are not definitions if they provide disjunctive lists which are either closed or open to future revision

The two types of theory are not the same, so evidence of the venerability of the one is not automatically evidence of the venerability of the other

Of course, the theories might be linked In developing theories of the arts, the early moderns might have drawn implicitly on theories of art They might have reasoned

(1) P is an an art form if and only if the works of P are works of art
(2) x is a work of art if and only if x is φ
(3) so P is an art form if and only if the works of P are φ

Here, (2) states a theory of art in the work-oriented sense which delivers (3), a theory of the arts in the activity-oriented sense This is exactly how contemporary philosophers think up theories of the arts[14]

Did the early moderns reason on the model of (1) to (3)? The absence of any explicit statement like (2) is inconclusive it might be implicit Better evidence lies in the kinds of counterexamples the early moderns considered For example, Charles Batteux proposed that the arts imitate beauty in nature, and versions of the imitation theory gained dominance for several decades (*op cit*) When critics pointed out that music need not imitate beauty in nature, theorists spoke to the counterexample[15] Given that the imitation theory was tested against counterexamples, it is striking that nobody worried that Homer's battle scenes and Grunewald's *Crucifixion* are works of art but do not imitate beauty in nature Counterexamples to (3) come in for debate, but not counterexamples to (2) The best explanation of this is that the early modern imitation theory is a theory of the arts that does not imply an imitation theory of art That an art aims to imitate natural beauty does not imply that every work in the art is an imitation of natural beauty

Kristeller's history ends with Kant, and it is interesting to consider his art-theoretical concerns[16] In section 43 of the third Critique, Kant distinguishes art "as human skill" first from nature, then from science, and finally from craft In the next section, he distinguishes the agreeable arts that aim at mere enjoyment from fine or *schon* art that furthers those mental powers that facilitate social communication Sections 51 and 52 contain an extended discussion of each of the arts and their value For Kant, there are connections between genius and the fine arts and also between the fine arts and beauty, but he never attempts to define the fine arts with reference to a theory of art.

---

[14] For example, Dickie, *The Art Circle*

[15] Kivy, *Philosophies of Arts*, chapter 1

[16] Kant, *Critique of the Power of Judgement*, Paul Guyer and Eric Matthews, trans (New York Cambridge, 2000)

Hegel also inherited the early modern project.[17] He writes that "the real world of art is the system of the individual arts" (*ibid.*, p. 614), which function to "find for the spirit of a people the artistic expression corresponding to it" (*ibid.*, p. 603), where an artistic expression is in a sensuous medium (*ibid.*, p. 621). Thus sculpture expressed the spirit of the Greeks and music that of the Germans.

Other nineteenth-century writers have other fish to fry. Many speculate about the elements of greatness in art and hence (as they see it) the character of artistic genius, a Kantian theme. By taking what is great in art to be the work of the genius, these writers inch towards theories of art. Indeed, the search for a theory of art requires a grip on the idea that the products of the arts are works—an idea that may only have developed at this time.[18] At any rate, recent philosophers are not interested in art defined as an expression of genius. For them, art includes Bunyan as well as Milton, Boilly as well as Poussin, and Mahavishnu Orchestra as well as Miles Davis.

The upheavals of the last century launch the contemporary project of defining art. The historian Michael Fried writes that the avant-garde "called into question . . . the already somewhat dubious concept of a 'work of art'" for it "equipped one to treat virtually any object as a work of art."[19] Turning to the philosophers, Beardsley confesses that his search for a definition of art sprung from "the enormous and even ridiculous variety of objects, events, situations, texts, thoughts, performances, reframings from any performance, and so on that have, in recent times, drawn the label 'artwork' from their authors, admirers, or patient endurers."[20] Likewise, George Dickie pays due to "the strange and startling innovations of Duchamp and his latter-day followers" as the inspiration for his theory of art.[21] The Kristeller of our times is Noel Carroll:

> it is in the twentieth century that the theoretical task of coming to terms with virtually continuous revolutions in artistic practice has become urgent. That is, it is in the twentieth century that the problem of identifying art has become persistently unavoidable. . . . The recurring task of

---

[17] Hegel, *Aesthetics: Lectures on the Fine Arts*, T. M. Knox, trans. (New York: Oxford, 1975).

[18] Lydia Goehr, *The Imaginary Museum of Musical Works* (New York: Oxford, 1995).

[19] Fried, *Three American Painters* (Cambridge: Harvard, 1965), p. 47.

[20] Beardsley, "Redefining Art," p. 298.

[21] Dickie, "The Institutional Conception of Art," in Benjamin Tilghman, ed., *Language and Aesthetics* (Lawrence: Kansas UP, 1973), pp. 21–30, here p. 27.

> the philosophy of art has been to provide means to identify new and emerging work, particularly work of a revolutionary sort, as art[22]

Whereas Batteux and Diderot seek a theory of the arts, Beardsley and Dickie seek a theory of art

Consider Weitz's landmark anti-essentialism (*op cit*) Weitz argues that there is no correct definition of art, where a definition of art states individually necessary and jointly sufficient conditions for an item's being a work of art Weitz assumes that art cannot be defined if the concept of art is an open concept An open concept, first, has application conditions that are in principle emendable and, second, is not closed by stipulation The application conditions of the concept of art are emendable because new cases "have constantly arisen and will undoubtedly arise which will demand decisions on the part of those interested, usually professional critics, as to whether the concept should be extended or not" (*ibid*, p 32) Moreover, the concept cannot be closed by stipulation since that "forecloses on the very conditions of creativity in the arts" (*ibid*, p 32) Therefore art cannot be defined Commentators quickly complained that artistic creativity is not incompatible with the concept of art being closed, unless open concepts question-beggingly entail creativity[23] Be that as it may, the argument also implies the empirical claim that tomorrow's avant-garde artworks falsify today's definitions of art Weitz targets definitional theories of art, not theories of the arts, and what drives his anti-essentialism is avant-garde art production

A second landmark argument also bears out Carroll's observation that the contemporary concern is with theories of art In his 1964 paper "The Artworld," Arthur Danto argues as follows[24] If art is defined in respect of perceptible features alone, then no artwork is perceptually indiscernible from a nonartwork But some artworks are perceptually indiscernible from nonartworks So art is not defined only in respect of perceptible features The situation promotes a specific kind of error Danto imagines a fictional Mr Testadura who visits his local MoMA and cannot tell the art from the nonart—he mistakes Robert Rauschenberg's *Bed* for the kind of item sold at Sears Again, what drives this argument are works like *Bed* that are perceptually indiscernible from "real objects"—and that the Testaduras of the world mistake for nonart

---

[22] Carroll, "Historical Narratives and the Philosophy of Art," *Journal of Aesthetics and Art Criticism*, LI (1993) 313–26, pp 313–14

[23] Maurice Mandelbaum, "Family Resemblances and Generalizations Concerning the Arts," *American Philosophical Quarterly*, II (1965) 219–28 See also Kivy, *Philosophies of Arts*, pp 36–37

[24] Danto, "The Artworld," *Journal of Philosophy*, LXI (1964) 571–84

A case can be made for (R) if we expect theories of art to address the challenges of the avant-garde and if, in addition, theories of art are not in fact needed to address these challenges. We do indeed expect theories of art to address the challenges of the avant-garde. Are we wrong to do so?

II. THE STATE OF THE ART

Philosophy often obeys Newton's third law of motion. Fifty years ago Weitz tried to scotch the project of definition forever (*op. cit.*),[25] but his efforts led instead to an explosion of new theories of art, most of them definitions—a recent bibliography runs to more than four hundred items.[26] This activity has not been wasted: we know more about art than we did before. However, we now face an impasse, a deep disagreement about avant-garde works like Rauschenberg's *Bed*. The impasse does not impeach the new theories, but it warrants doing as (R) recommends and looking elsewhere for tools to cope with the avant-garde.

The standard typology divides theories of art into functionalist and proceduralist varieties.[27] On functionalist theories, what makes something a work of art is a function it has. Poisons are defined functionally, since what makes something a poison is its having the function of causing harm when absorbed into the body. Recent functional theories of art assign works of art an aesthetic function such as "affording an experience with marked aesthetic character."[28] On proceduralist theories, what makes an item an artwork is a matter of the "procedures, rules, formulas, recipes, or whatever by which artworks are generated."[29] Knighthood is defined procedurally: Sir Mick is made a knight by the Crown's conferring the status upon him. In the case of artworks, the "whatever" may be an act of someone who is authorized within an informal institution to confer art status.[30]

The dispute between functionalism and proceduralism is not a dispute about whether artworks have a function. A proceduralist might argue that since nothing is a function of all and only artworks, functionalism is false, so proceduralism is true. However, that is just one route to proceduralism Procedural theories of art are consistent with works of art having a function By the same token, functionalism is consistent with there being procedures for generating all and only artworks A functionalist might argue that since there are no procedures for generating artworks, proceduralism is false, so a functional theory is true, or she might just as well reject the opening premise of this argument

---

[25] See also Paul Ziff, "The Task of Defining a Work of Art," *Philosophical Review*, LXII (1953): 58–78, and William Kennick, "Does Traditional Aesthetics Rest on a Mistake?," *Mind*, LXVII (1958): 317–34.

[26] Jinhee Choi, "Bibliography," in *Theories of Art Today*, pp. 241–62.

[27] Davies, *Definitions of Art*.

[28] Beardsley, "Redefining Art," p. 299. See also James Anderson, "Aesthetic Concepts of Art," in *Theories of Art Today*, pp. 65–92, and Iseminger, *The Aesthetic Function of Art*.

[29] Davies, *Definitions of Art*, p. 30.

[30] Dickie, *The Art Circle*.

Functional and procedural theories render contrary verdicts when there is a gap between function and procedure—they represent a disagreement over puzzle cases Some puzzle cases are imagined,[31] but plenty are real, though a few get recycled in the literature Marcel Duchamp's *Fountain* of 1917 is a urinal from a plumbing supply store Andy Warhol's 1964 *Brillo Boxes* are plywood boxes painted to match Brillo soap pad boxes stocked at grocery stores In 1953, Robert Rauschenberg erased a drawing by Willem de Kooning and showed it as the *Erased de Kooning Drawing* A year before, Rauschenberg's friend, John Cage, had premiered *4'33"*, a work "for piano " Functionalists reason that unless these works have the function that makes for artworks, they are not art Proceduralists counter that they are art if they are generated by the required procedures Dickie claims as a virtue of his procedural definition that it counts the puzzle cases as art, Beardsley claims as a virtue of his functional definition that it withholds them status as art [32]

As long as our goal is to cope with the puzzle cases, we seem to have two options We can ride with functionalists, who stick to their guns and deny that the puzzle cases are artworks Or we can ride with the proceduralists and carry the flag for the puzzle cases We must decide which class of theory gets the extension of art right

This would not be the first time in philosophy that a choice between theories represents a clash in intuitions A dilemma is not by itself an impasse—it might be the starting point for making progress Theories of art might rectify intuitions when there are independent grounds for theory choice—grounds that go beyond extensional adequacy

No theory is worth choosing unless it has content, coherence, and internal consistency Critics fault versions of functionalism and proceduralism by all three standards For example, some say that Dickie's institutional theory of art is viciously circular, hence without content [33] Functionalism faces a similar charge Beardsley claims that all and only works of art afford experiences with a marked aesthetic character: one gets the experience from Bob Marley's "Duppy Conqueror," T. S. Eliot's "Ash Wednesday," and Martha Schwartz's *Bagel Garden*, but not from Olympic hockey, the essays of David Lewis, or a Macy's window display. The worry is that no notion of an experience with marked aesthetic character both gets this sorting right and can be characterized independently of the sorting. Gallons of ink have been spilled making and answering these charges. Let us acquit the accused just to see where we get. Stipulate that versions of functionalism and proceduralism are contentful, coherent, and internally consistent.

---

[31] Davies, *Definitions of Art*, pp 100–06

[32] Beardsley, "An Aesthetic Definition of Art," p 25

[33] Wollheim, *Painting as an Art*, Gaut, "'Art' as a Cluster Concept "

Apart from the general virtues of any theory worth having, some propose a trait that specifically marks a good theory of art.[34] Art has value as a whole and a theory of art might explain art's value. Moreover, individual artworks have value and a theory of art might ground critical principles that determine what value accrues to any given artwork. Thus one might prefer a theory of art if it explains the value of art as a whole or determines the principles attributing value to individual works of art.

The trouble is that this criterion of theory choice is not common ground among functionalists and proceduralists. It favors functionalism and functionalists favor it; it disfavors proceduralism and proceduralists reject it in turn. This is explicit in exchanges between Beardsley and Dickie. Insisting that a theory of art should solve the mystery of art's value, Beardsley notes that his own theory implies that "it is good for us to experience, at least occasionally and to a degree seldom made possible except by artworks, the immediate sense (say) of inclusive self-integration and complex harmony with phenomenal objects."[35] Dickie agrees that we need an explanation of the value of art as a whole and a grounding for critical principles, but he denies that an adequate theory of art must supply the explanation. He awards his own theory a green star for hiving off the evaluative tasks to a separate theory of art value.[36]

No general principles of theory choice decide this dispute between Beardsley and Dickie. Analogies with other domains cut both ways. It seems reasonable to define life in a manner that does not imply an account of the value of life. It also seems reasonable that a theory of an institution like money should explain the value of money. One suspects that it is intuitions about the extension of art that drive not only the choice of one class of theories over the other but also that drive the selection of criteria for theory choice

---

[34] Beardsley, "Redefining Art."

[35] Beardsley, "In Defense of Aesthetic Value," *Proceedings and Addresses of the American Philosophical Association*, LII (1979): 723–49, here p. 743.

[36] Dickie, *Evaluating Art* (Philadelphia: Temple, 1988).

If the suspicion is accurate, then recent art theorists endorse incompatible types of theories because they rely (1) on different intuitions about the puzzle cases and (2) on different criteria of theory choice, where (2) is also determined by (1) That is an genuine impasse The impasse does not impeach functionalist or proceduralist theories of art—it is methodological, not probative But it brings up this question What if we do not look to theories of art to help cope with the puzzle cases?

III WHAT WE NEED

History distinguishes theories of the arts, which were pursued by the early moderns, from theories of art, which are now used to address the challenges of the avant-garde But why look to theories of art to address these challenges? Why not theories of the arts instead? Or theories of the art forms—that is, theories of music, dance, theater, literature, film, painting, architecture, and the like We have three types of theory to choose from

theory of art item *x* is a work of art if and only if
theory of the arts activity *P* is an art form if and only if
theory of an art form item *x* is a work of art form *P* if and only if

If (R) is our theory of art, then we need theories of the arts or the art forms to cope with the avant-garde

Carroll observes that "avant-garde mutations often strike the public and some critics as unintelligible, and therefore, as not art "[37] The scenario should ring a bell The public is presented with an item—a urinal or a shark in a tank of formaldehyde—which defies expectations and induces bafflement Sometimes the bafflement feeds a desire to know more, which feeds a cottage industry of slim primers on the nature of art Sometimes it feeds a suspicion that we are victims of a hoax Theories of art are crafted to address the bafflement

This bafflement can take one or both of two forms On one hand, it might be conceptual That is, avant-garde works might confound a concept—a concept of art, a concept of the arts, or a concept of a specific art form On the other hand, the bafflement might be theoretical Avant-garde works might put pressure on a theory of art, a theory of the arts, or a theory of a specific art form The distinction between the two kinds of bafflement is important It is possible that to have a concept of art or the arts or an art form is to know a theory of art or the arts or the art form However, we cannot assume that having any one of these concepts consists in knowing the corresponding theory Just as a person can have a concept of water without knowing that it is H_2O, perhaps a person can have a concept of art without knowing what makes an item a work of art This would mean only that having the concept involves deference one may have a concept of art and defer to the experts about the nature of art

---

[37] Carroll, "Historical Narratives and the Philosophy of Art," p 316

Begin with conceptual bafflement Suppose that to have a concept of art is, at a minimum, to be pretty reliable at sorting items that are works of art from items that are not works of art William Kennick set a test for having this concept we have a concept of art just in case we mostly succeed when sent into a warehouse under orders to bring out all and only works of art (*op cit*, pp 321–22) Mr Testadura fails Kennick's test Sent into the warehouse, he mistakes art for nonart he leaves behind avant-garde works like *Fountain*, *Brillo Boxes*, and *4′33″* And we are all Testaduras Avant-garde works confound our ability to identify works of art, showing us to be poor art detectors Hence our bafflement

In point of fact, we are not one bit like Testadura In normal contexts, anyone minimally schooled in art can reliably tell art from nonart Nobody normally mistakes Rauschenberg's *Bed* for an ordinary bed or *4′33″* for an unscheduled intermission That is, nobody has any trouble identifying artworks in museums or concert halls, set atop pedestals, or performed by people sitting at pianos In normal conditions, *Fountain* and *4′33″* are no less markedly artworks than *Winged Victory* or *The Magic Flute* The "normally" is obviously crucial The capacity to identify an item of a kind often requires training and depends on context I can tell red from orange in normal conditions, not any conditions I can identify my keys in part because I know what they look like and also because they are in the dish in the front hall Placed in context, the urinal or four and a half minutes of not playing the piano are blatantly works of art Special conditions, like Kennick's warehouse, do subvert art recognition, but that is not surprising Many concepts are confounded in some conditions

Two facts, one about recent art and one about our concepts, drive the point home First, it is possible to create artworks whose identification requires picking up on contextual cues, and a great deal of recent art plays on context in this way As an art critic observes, although much recent art "is rooted in an anti-museum attitude it is museums and art galleries that are the primary locus for such art [which] needs an institutional context to be seen "[38] Second, people often defer to experts who know what makes an item a work of art Moreover, the two points are connected Such facts about an item as its having been placed on a pedestal trigger deference by indicating the experts' verdict

---

[38] Michael Rush, *New Media in Late Twentieth Century Art* (New York Thames and Hudson, 1999), p 116

Concepts of art forms such as music, dance, and painting are not in quite the same boat as the concept of art First, context seems to play a lesser role in identifying items in at least some art forms Second, most people defer less to experts on what items count as works of music, dance, or painting, for example

If so, then perhaps bafflement at the avant-garde springs not from a challenge to the concept of art but instead from a challenge to art form concepts Music is patterned sound and we recognize works as works of music by hearing patterned sound Nobody attending a performance of *4′33″* doubts that they are in the presence of art That is what they know for sure and will bet money on The trouble is that they are told that they are listening to music and yet they identify music by listening for patterned sound Likewise, nobody takes *Erased de Kooning Drawing* for a mere sheet of paper for it is displayed as art What is baffling is how it is a drawing, for it does not fit the concept of a drawing

Conceptual bafflement has an evaluative side artworks are made to be appreciated, not identified So if Testadura does not exist, his slightly more sophisticated cousins do They recognize artworks and yet are baffled because they do not know how to appreciate them They are apt to remark that a small child could do the same The usual explanation of their bafflement is that they do not employ up-to-date principles of art criticism They judge *Fountain* by the same lights as *Winged Victory* and *4′33″* by the same lights as a Chopin étude

Here is another hypothesis Testadura's cousins do not need principles of art criticism, because they need critical principles pertaining to an art form Kennick remarks that "only a man corrupted by aesthetics would think of judging a work of art as a work of art in general, as opposed to this poem, that picture, or this symphony" (*op cit*, p 329) Testadura's cousins recognize *Erased de Kooning Drawing* as pictorial art, but remain baffled Having attended Richard Wollheim's Mellon lectures,[39] they appreciate pictures as surfaces marked up so as to sustain experiences of seeing-in, experiences that reveal what the markings meant to the artists who made them *Erased de Kooning Drawing* yields nothing along these lines, for there is nothing to see in it The hypothesis is that avant-garde works challenge traditional principles of music appreciation, picture appreciation, dance appreciation, and the like. They do not challenge generic principles applying equally to *Middlemarch*, "My Aim Is True," *Brillo Boxes*, and *Sagrada Familia*.

---

[39] Wollheim, *Painting as an Art*

None of this plays down the challenge of the avant-garde. Avant-garde works are baffling. However, the bafflement stems from a challenge to art form concepts rather than the concept of art, or to art-form specific principles of criticism rather than generic principles of criticism.

Turn next to theoretical bafflement. As we saw, having a concept of art (or an art form) need not amount to knowing a theory of art (or the art form). Most people have no trouble identifying art, but they defer to the experts about the nature of art. Still, wondering what the experts know can induce bafflement. On the standard story, we find ourselves in Kennick's warehouse and exclaim, "Ugh, which of these is art?" This story is fiction. Still, we might truly find ourselves in the MoMA and exclaim, "This is art! Ugh, what are these people thinking?" We want to know what theories the experts wield—the theories knowledge of which makes them experts. Such bafflement is theoretical rather than conceptual.

Theoretical bafflement is bafflement that is dispelled by revising one's theories. Grant that works like *Fountain* or *4′33″* induce theoretical bafflement. The question is what kind of theory is needed to address the bafflement. According to the standard story, we need a theory of art. We wonder: If *Fountain* or *4′33″* is art then what in blazes makes anything art? We look to a theory of art to settle the question. Some adopt procedural theories that rule in *Fountain* and *4′33″* and others turn to functional theories that rule them out. But maybe all we need is a theory of the arts. We should wonder: If *Fountain* is sculpture then what in heck makes anything sculpture? Or if *4′33″* is music then what, pray tell, makes anything music? Theories of sculpture and music settle these questions.

The two kinds of questions and the kinds of theories that settle them are asymmetric. Take *Fountain*. There is nothing left to settle when a theory of sculpture settles the status of *Fountain* as sculpture. A theory of sculpture that settles whether *Fountain* is a sculpture leaves nothing to be settled by asking whether *Fountain* is art. The converse does not hold. A theory of art that settles whether *Fountain* is art leaves unsettled the question whether it is sculpture. Perhaps it belongs to a new art form like conceptual art.[40]

---

[40] Lopes, "Conceptual Art Is Not What It Seems," in Peter Goldie and Elizabeth Schellekens, eds., *Conceptual Art and Philosophy* (New York: Oxford, 2007), pp. 238–56.

Thus we do not need a theory of art except to address cases where an answer to the art form question leaves open the answer to the art question We need a theory of art only if it is settled by a theory of each art form whether a work belongs to the art form and yet it is not settled whether the work is art

In particular, we do not need a theory of art to cope with the philosophers' puzzle cases It is not settled by a theory of sculpture that *Fountain* is sculpture or by a theory of music that *4′33″* is music What about other cases? It is not settled by a theory of landscape architecture that Christo's wraps are architecture, by a theory of painting that Ben Vautier's "Regardez moi cela suffit je suis art" is painting, by a theory of theater that a performance modeled on Augusto Boal's "theater of the invisible" is theater, or by a theory of poetry that bp nichols's "Cold Mountain" is poetry We need a theory of art to cope with the avant-garde only if there are works that contribute to our bafflement whose art forms are settled while their status as art is open

There are no such works as long as every work of art belongs to at least one art form The theoretical challenge posed by any given work can be settled in four ways The first finds that a theory of an art form admits the work and it is an artwork The second finds that no theory of any art form admits the work, but it is not an artwork The third finds that no theory of any art form admits the work, but it is an artwork The fourth finds that no theory of any familiar art form admits the work, but the work pioneers a new art form Only the third case shows that we need a theory of art However, any reason we have to say that a work is an artwork that does not belong to any art form is reason to say that it pioneers an unfamiliar art form The move is not ad hoc It represents how things really stand It is a fact about the arts that their number grows—notable recent additions being photography, film, video, computer art, and conceptual art That a work belongs to no familiar art form and yet is an artwork is reason enough to project a new art form

To sum up, the avant-garde is baffling conceptually and theoretically A case can be made that the conceptual bafflement is bafflement about art form concepts—so long as we rely less on context and defer less to experts in applying art form concepts than the generic concept of art Moreover, we need a theory of art to address the avant-garde only if theories of the arts leave it some work to do The avant-garde works that baffle philosophers do not meet this condition

If attachment to traditional theories of art is hard to shake, then maybe the culprit is a category mistake Each art form is obviously a topic of study which implicitly embeds a theory of that art form, so we have theories of the arts Then, just as we wonder what men and women make up the division apart from the battalions, batteries, and squadrons, we wonder what is art apart from each of the arts[41] The mistake is to treat art like another one of the arts—to look for a theory of art to place alongside theories of music, pictures, drama, and the rest

IV WHAT WE CAN HAVE

If the avant-garde is baffling, then what is the right description of the bafflement? The assumption has been for several decades that it only makes sense to describe the bafflement as a challenge to concepts or theories of art But maybe the bafflement is a challenge to concepts and theories of the art forms—or some of them, anyway How do we choose between the two descriptions? Intuitions can hardly be expected to respect the required distinctions between concepts and theories of art and the art forms We must choose the option that helps us best understand what we need to understand As long as we do not need a theory of art to address the avant-garde, we can have (R) (R) is no less informative than it need be It passes the buck to theories of the individual art forms and theories of the arts as a whole, and it is from these that we should expect answers to our questions Of course, the prospects for these theories had better look good In fact, their prospects improve if (R) is true

The buck passes first to theories of the art forms These tell us what makes some items but not other items works in a given art form

Here is a hurdle Painting is one of the arts If a work of painting is a product of that activity, then a dirty rag is a work of painting But it is not an artwork The reply is that works in an activity are not merely products of the activity, some activities have products only some of which are works A start is this *x* is a work in *P* just in case *x* is a primary product of *P*—that is, a product of a kind upon which the existence of *P* conceptually depends A dirty rag is not a primary product of painting because there could be painting without dirty rags Since there can be no painting without paintings, paintings are primary products of painting So paintings are works of painting

Another objection targets theories of the arts conjoined with (R) Grant that every work of painting, music, and theater is a work of art—that is contentious enough Surely some pots are works of art, but not the ones I buy from Walmart, and some buildings are works of art, but not my garden shed. And if ceramics and architecture are arts, and pots are ceramics and sheds are buildings, then any pot or shed is an artwork. Better reject (R) than accept such an absurd result.

---

[41] Gilbert Ryle, *The Concept of Mind* (London Hutchinson, 1949), pp 16–17

The reply displays some advantages of (R). Nothing requires the identification of works in an art form with works in a medium like ceramics or building. There is a distinction between ceramics and works in the ceramic art form, between buildings and works of architecture. However, the distinction can be made differently for different arts. Works of architecture may have a function that mere buildings lack, and works in the ceramic art form may be generated in accordance with special procedures (maybe they are made by MFAs). Traditional theories of art require the very same distinction, but attempt to make it uniformly—so that what distinguishes *Sagrada Familia* from a garden shed also distinguishes bizen ware from Walmart ware. Theories of the arts conjoined with (R) need not take a uniform approach. First, functional theories of music and painting might assign to each a different function. Maybe paintings are imitative and musical works are expressive. Second, it is possible that some art forms are best characterized functionally and others procedurally.

Moreover, functionalist theories of the art forms are more promising than functionalist theories of art. The latter posits a function shared by "Duppy Conqueror" and "Ash Wednesday," but not by a Rangers game or a window display at Macy's. Dance has it, but not ice dance, ballet has it, but not gymnastics, novels have it, but not biographies. By contrast, a functional theory of music may characterize the function of musical works as specific to music. Think of Eduard Hanslick's account of a "specifically musical kind of beauty" in "tonally moving forms."[42] Thinking of Hanslick suggests where to look for the function, if there is one, for it will be implicit in musicology, music theory, music criticism, and the psychology of music. Functional theories of the arts are in development,[43] and so are theories of the aesthetic specific to art forms.[44]

Finally, functional theories of the arts are less vulnerable to avant-garde counterexamples than functional theories of art. Suppose that a functional theory of theater implies that a work of "invisible theater" is not theater. This result is not an impasse: it is a fruitful discovery. It does not follow that the work is not art, and if the work is art then it follows from that fact together with (R) that the work is a work in another, perhaps unfamiliar, art form The innovations of the avant-garde are represented on this view as innovations

---

[42] Quoted in Kivy, *Philosophies of Arts*, p. 28.

[43] For example, Wollheim, *Painting as an Art*; Peter Lamarque and Stein Haugom Olsen, *Truth, Fiction and Literature* (New York: Oxford, 1994), and James Hamilton, *The Art of Theater* (Malden, MA: Blackwell, 2007).

[44] For example, Lopes, *Sight and Sensibility: Evaluating Pictures* (New York: Oxford, 2005).

The buck also passes to theories of the arts as a whole Not all activities are arts Painting, poetry, and architecture are arts, window dressing, philosophy, and garden shed fabrication are not Which activities are arts? This is the question, now roundly ignored, that troubled the early moderns

They sought functional theories of the arts Kristeller himself was skeptical, remarking that "the system of the fine arts is hardly more than a postulate and most of its theories are abstracted from particular arts, usually poetry, and more or less inapplicable to the others" (*op cit*, p 46) Many will agree Do painting, video art, landscape architecture, and literary prose writing have any function that is not a function of tattooing, fly tying, gymnastics, philosophy, and perfumery? Why count dance and calligraphy as arts but not ice dance and paper-making, or mime but not clowning, sculpture but not body building, opera but not World Cup soccer? Skeptics about functional theories of the arts have an option not open to the early moderns—procedural theories Nobody has developed such a theory, though it is worth trying

A fascinating, unexplored body of literature is ready at hand During the past hundred years, several new art forms have been recognized following debates about their admissibility Proponents of photographic art, conceptual art, and video art might have appealed to the same general criteria for any activity to be an art form If so, their reasoning might suggest a functional theory of the arts Alternatively, their reasoning might show that only a procedural theory of the arts has any plausibility Either way, (R) is consistent with functional and procedural theories of the arts

The two bucks are passed independently (R) allows that theories of the art forms can be developed without waiting for a theory of the arts And if there be worthy questions about why some activities are arts, we need not wait upon answers in developing theories of the art forms

A final worry One might accept all this buck passing and still reject (R) After all, (R) is implicitly disjunctive and so implies that there is no unified phenomenon of art That may be fine as a bit of philosophy, but the concept of art is deployed in theoretical contexts—in scientific and historical hypotheses, for instance—and we need a theory of art to make explicit the nature of art as an object of study outside philosophy Beardsley writes that "the definition of 'artwork' should be of the greatest possible utility to inquirers in other fields besides aesthetics" and he is surprised that nonphilosophers are "wont to spurn the friendly offers of aestheticians to help them out "[45]

The response of art scholars makes perfect sense if they make no use of a concept of art in their work on art (even if they use a concept of art in discussing our concept of art) Art is not the object of any study There are no serious psychological, anthropological, sociological, or historical hypotheses about art, although there are serious hypotheses about pictures, music, and the novel Moreover, there is no reason to expect such uniformity across studies of the arts as might imply a uniformity of theories Film historians need not think of their object of study in a way that is consistent with the way music theorists think of their object of study, for there are no hypotheses about movies and songs as a single phenomenon Or, if there are hypotheses about works in all of the arts they are also hypotheses about a lot of other phenomena besides[46] Beardsley almost sees this when he writes that "it stands to reason that someone starting out to write a history of (visual) art would want to have a reasonably definite idea of what it is he is writing a history of "[47] Drop the parentheses around "visual" and the claim is obvious, for a historian of visual art perforce deploys a concept of visual art It does not follow, however, that she employs a concept of art in any sense except that given in (R)

This is no accident if it reflects the fact that there is nothing left to explain when you have theories of the art forms and the arts

V CONCLUSION

Although avant-garde puzzle cases create demand for theories of art, they are best addressed by theories of the individual arts and theories of the arts as a whole All we need in a theory of art we get from (R) Happily, (R) breathes new life into functionalism, because functional theories of the arts stand a fighting chance It also tips a hat to proceduralism since procedural theories of the arts are especially plausible Perhaps, in the end, (R) is less radical than it seems Its purpose is to change the subject, but the new subject is one that is already getting attention from philosophers who are crafting theories of the arts True, we must also revive the question that troubled the early moderns, but having more work to do is a blessing unless it cannot be done

DOMINIC MCIVER LOPES

University of British Columbia

---

[45] Beardsley, "Redefining Art," p 304

[46] For example, Howard Gardner, *Art, Mind, and Brain* (New York Basic, 1982), and Ellen Winner, *Invented Worlds The Psychology of the Arts* (Cambridge Harvard, 1982)

[47] Beardsley, "Redefining Art," p 304

WHAT COLORS COULD NOT BE AN ARGUMENT FOR COLOR PRIMITIVISM*

In "What Numbers Could Not Be," Paul Benacerraf argues, famously, that numbers cannot be what many traditional logicists had hoped that they would turn out to be particular sets[1] The essence of the problem is that there are many distinct systems of sets that form progressions, and that forming a progression is all that a system of sets needs to do to serve as a basis for arithmetic This paper offers a related argument for primitivism in the philosophy of color Colors are irreducible properties of objects properties that are presented in visual experience, and that stand in certain similarity relations in virtue of how they seem—or, to be more precise, in virtue of how they make objects that have them seem There is no need to identify colors with any of the various physical or dispositional properties with which they happen, actually, to covary Such identifications will end up endowing colors with properties they do not have and will commit us to the truth of counterfactuals to which we would do better to remain uncommitted The conclusion is that the dispute between the various varieties of reductive color realism rests on a mistake

I STAGE SETTING WHY NUMBERS ARE NOT SETS

The problem for a Fregean account of the numbers—an account that tries to construct them from the materials of some form of set theory—is that there are many progressions that can be so constructed That is, there are many distinct systems of sets such that it is possible to pick out one set that functions as a zero, and to specify a successor relation between sets that guarantees a unique successor for each set One such system might use the sets $\{\varnothing\}$, $\{\{\varnothing\}\}$, and $\{\{\{\varnothing\}\}\}$ to represent the numbers 1 through 3, while another might use the sets $\{\varnothing\}$, $\{\{\varnothing\}, \varnothing\}$, and $\{\{\{\varnothing\}, \varnothing\}, \{\{\varnothing\}\}, \{\varnothing\}, \varnothing\}$[2] The abundance of such progressions is responsible for two embarrassments for those who wish to identify, in a strong way, the natural numbers with specific sets The first embarrassment is that any choice between rival pro-

---

*Thanks to Michael Watkins and to Jonathan Cohen for comments on an earlier version

[1] Benacerraf, "What Numbers Could Not Be," in Hilary Putnam and Benacerraf, eds, *Philosophy of Mathematics Selected Readings* (New York Cambridge, 1983), pp 272–93

[2] The definition of the successor relation for the first of these progressions is obvious For the second, it is '*x* is the power set of *y*'

0022-362X/08/0503/128–55

gressions seems completely arbitrary. What could ground the truth of the claim that this progression, rather than that one is the "true" series of natural numbers? The second embarrassment is that, even after one has managed to convince oneself that one has identified a particular series of sets with the natural numbers, one will still be forced to give sense to certain questions about the numbers that do not seem very plausibly to have sense. For example, on the first progression given above, it turns out that 3 contains 2, but does not contain 1. On the second, however, 3 contains both 2 and 1. Does 3 contain 1, then, or does it not? In fact the correct answer to this question is that it is ill-posed, despite receiving two distinct and clear answers on the two systems described. One way of putting this problem is that if numbers are not actually sets then any view that wrongly identifies them with sets will also, at the same time, endow them with properties that they do not have.[3]

If numbers cannot be identified with sets, it might seem that they will need to be identified with some other sort of entity. A moment's reflection, however, will reveal that precisely the same embarrassments lie in store for anyone who pursues this sort of line. But if numbers cannot be identified with any things at all, how can it be that there are any arithmetical truths? Benacerraf's answer is Wittgensteinian in flavor. We learn to produce the counting numbers much as we learn the sequence of the alphabet, although in the case of the numbers we also learn a recursive procedure for generating them. Moreover, we also use these numbers to count collections of things, in order to answer questions about how many things there are in those collections. These two uses of number words yield a collection of truths about numbers (two comes after three, if a group contains five things then it contains more than three things) without any need to suppose that the number words are names for anything at all or at any rate, for anything other than, precisely, numbers.

II. THE CASE OF COLOR

One might think that it would be hard to extend Benacerraf's points about the numbers to the case of colors. Numbers, after all, are abstract entities if they are entities at all. Colors, on the other hand, seem to be properties of physical objects, and some philosophers have thought that they were in fact physical properties[4] But the conclusion of Benacerraf's argument can be expressed in the following slogan Numbers form an autonomous domain[5] This slogan is meant to suggest that arithmetical truths do not need to be explained in terms of anything other than the most basic truths about numbers truths we learn when we learn to count out loud, and when we learn to count out how many objects are in a collection The primitivist about color is making a similar claim the central truths about color do not need to be explained in terms of anything outside of the domain of color Of course this claim might be true even if some sort of *ontologically* reductive account were true, since we can separate out ontological and explanatory issues[6] But the color primitivist also holds that the reason why we cannot explain these central truths in a reductive way is that, ontologically, colors are primitive

---

[3] Benacerraf's argument has, of course, been disputed. See, for example, N. P. White, "What Numbers Are," *Synthese*, xxvii (May-June 1974): 111–24. But no one disputes that it has a great deal of initial plausibility. Primitivism about color, on the other hand, is decidedly a minority view. It is the point of this paper to increase its plausibility. For this reason, taking Benacerraf's point as given should be regarded as dialectically permissible.

The claim that central truths about color do not need to be explained in terms of anything outside of the domain of color will initially strike many—especially advocates of reductive realist views of color—as quite obviously false Are not various facts about reflectance properties, and the neurophysiology of the human visual system, crucial for a proper understanding of color? Are they not therefore facts that are outside the domain of color, but that explain truths about color? The answer to this challenge is It depends on what one means by 'a proper understanding of color' Normal people can certainly have color concepts without having any idea about the propagation of light Nevertheless, a color scientist who does not know that there are three different color receptors in the normal human retina does not have a proper understanding of color—for a color scientist But even before one becomes a color scientist, one typically has a complete mastery of color concepts It is this sort of mastery that I am claiming involves the central truths about color, and does not depend on explanations offered in terms of anything outside of the domain of color Of course, primitivism about color does not involve the denial that there is much that science can teach us about the causes of color experience Science might even teach us that nothing actually has the color properties presented to us in experience [7] The question is not whether science can teach us these things, but whether one of the things it can teach us is which properties, specified without the use of color terms, the colors actually turn out to be The primitivist claims it cannot teach us this, because colors simply are not properties that can be so specified [8]

---

[4] That *P* is a property of a physical object does not imply, of course, that *P* is a physical property A mountain, for example, can be beautiful And property dualists will hold that brains can have mental properties that are not physical properties

[5] In this connection, it is worth noting the title of one primitivist manifesto Justin Broackes, "The Autonomy of Color," in Alex Byrne and David Hilbert, eds , *Readings on Color*, Volume 1 (Cambridge MIT, 1997), pp 191–225

[6] Such a separation is what allows for the positions advocated by Donald Davidson in his "Mental Events," in his *Essays on Actions and Events* (New York Oxford, 1980), pp 207–25, and Jerry Fodor in his "Special Sciences," *Synthese*, xxviii (October 1974) 97–115

It should be admitted immediately that the fact that people can have complete mastery of color concepts without knowing anything about the wave theory of light does not by itself show much, one way or the other, about whether the colors might be properly identified with physical or dispositional properties of surfaces, or properties of light After all, it is possible to have complete mastery of the concept of water without knowing anything about chemistry Nevertheless it remains exceedingly plausible that water is properly identified with H_2O So, too, for all that has been argued so far, it might be that redness is properly identified with the possession of a spectral reflectance curve that falls within a certain set However, if every suggestion as to the nature of redness should turn out to offer us a property that stands to redness in much the same way as the various sets mentioned above stand to the number 3, then it will be plausible to conclude that we need not—and probably should not—select any particular property at all as actually *being* redness Rather, some other relation will obtain between the various candidate properties and redness For example, those properties will turn out to be co-instantiated with the colors, and may turn out to share certain structural features with the colors But such relations amount to something far short of identity

There are, then, two models in terms of which we might understand the relation between color properties and the properties that reductive realists have identified with them. One model, advocated in this paper, makes the relation similar to that which obtains between the number 3 and the sets $\{\{\{\emptyset\}\}\}$ and $\{\{\{\emptyset\}, \emptyset\}, \{\{\emptyset\}\}, \{\emptyset\}, \emptyset\}$. The other model makes the relation similar to that which obtains between water and H_2O. How can we decide between these models? It might be useful to begin by pointing out one salient difference between the two models. Science has no alternative suggestion as to what water might be, if it is not H_2O. On the other hand, the two sets mentioned as candidates for being the number 3 are on an equal footing. Moreover, there are many other rival candidates. This difference between water and numbers at least favors the arithmetical model as a model for color. For there are a number of candidates for the property of redness that fill that role fairly well. Redness might be a surface spectral productance. Or it might be the highly disjunctive physical basis for that productance. Or it might be a disposition to stimulate the longer-wavelength receptors in a human eye to a degree greater than the middle-wavelength receptors. Or it might be a disposition to produce certain experiences in human beings under certain conditions.[9] It is quite clear that these properties will yield coextensive classes of colored objects, at least in the actual world, and at least if it makes sense to talk about the extensions of color terms at all. After all, any object that has a surface spectral productance appropriate to redness will also have the disjunctive physical basis for that productance. Similarly, in determining that red things stimulate the longer-wavelength receptors to a greater degree than the middle-wavelength receptors, the criterion that has been used is that this ratio of stimulation is what produces a certain sort of experience. Thus this proposal will yield extensions for the color terms that are coextensive with those yielded by the fourth proposal. A similar argument goes for the first and third proposal, so that all four proposals can be seen to be extensionally equivalent.

---

[7] It might do this by showing that individual variation in color experience is so great that no sense can be attached to the claim that a specific object actually has a color C L Hardin pushes this line very hard in many places See, for example, his *Color for Philosophers Unweaving the Rainbow*, expanded edition (Indianapolis Hackett, 1993), pp 76–77 I argue against such a view, and also against an overreaching color realism, in "A Realistic Colour Realism," *Australasian Journal of Philosophy*, LXXXIV (December 2006) 565–89

[8] Of course in some sense color science *can* do this, since it can specify a particular color in the following way 'the color actual objects have when they have surface spectral reflectance so-and-so' But this specification, according to the color primitivist, fails to inform us of anything unless we already know what colors are So it fails in the task that reductive color realists take themselves to be engaged in telling us what colors are

On the other hand, there is a great deal of debate between advocates of various reductive realist positions. It must be admitted that the very existence of this debate represents an important disanalogy between reductive color realism and the position of a Fregean logicist

---

[9] The first two of these proposals were offered by, respectively, Alex Byrne and David Hilbert, "Color Realism and Color Science," *Behavioral and Brain Sciences*, XXVI (February 2003): 3–21, 52–63, and J. J. C. Smart, "On Some Criticisms of a Physicalist Theory of Colors," in Byrne and Hilbert, eds., pp. 1–10. The third is suggested by some remarks by Byrne in his "Color and Similarity," *Philosophy and Phenomenological Research*, LXVI (May 2003): 641–65. The fourth is the Lockean position, advocated by Christopher Peacocke in his "Colour Concepts and Colour Experience," in Byrne and Hilbert, eds., pp. 51–65, and by Colin McGinn in his *The Subjective View* (New York: Oxford, 1983). McGinn no longer holds this view.

But what the preceding considerations about extensional equivalence do suggest is that it will be almost impossible for one kind of color realist to show that another kind of realist has failed to capture the proper extension As a result, arguments between (for example) those who favor surface productances and those who favor the dispositional bases of those productances will turn on some other sort of consideration An examination of the disputes between realists reveals that very often these considerations consist in the claim that the advocate of a rival view is attributing to colors some features that we do not think colors actually possess And it is this sort of criticism that is exactly the sort that one logicist might level at another logicist who has offered a distinct proposal regarding which sets we should identify with the numbers Such criticism, unless one can avoid it oneself, is consistent with the idea that it is simply a mistake to identify colors with anything outside the domain of the colors That is, such criticisms, if they are both pervasive and generally correct, suggest that the characteristic error of nonprimitivist color realists is to identify colors with properties that themselves have or entail properties that we do not think colors possess And the only alternative realist view is that of the primitivist[10]

The following two sections of this paper fill out and support the above argument for color primitivism The first presents primitivism in a nonmysterious way The second shows that criticism of realist views does indeed often take the form suggested above, lending support to the idea that what is wrong with many reductive realist views may not be the realism, but the reductionism In order for the strategy to succeed completely, one would have to show that every reductive realist position ascribes too many properties to the colors That, obviously, is beyond the scope of this paper But I do hope to show that reductive views that fall into certain broad classes—primary-quality accounts and dispositionalist accounts—have a tendency to overascribe And I will argue that *any* sort of reductive view will give the colors too many modal properties

---

[10] If one is suspicious of primitivism, this may therefore lead one to some form of subjectivism or error-theory I give a broad argument for realism—an argument that does not favor a reductive or primitivist version of realism—in "A Realistic Colour Realism " For that reason I confine myself in this paper to arguments that presuppose some version of realism that validates the literal truth of such claims as 'grass is green' Thus I also do not discuss, in the present paper, versions of color realism that endorse a radical relativism about color properties For one such a view, see Jonathan Cohen, "Color Properties and Color Ascriptions A Relationalist Manifesto," *Philosophical Review,* cxiii (October 2004) 451–506

III. AN UNMYSTERIOUS PRIMITIVISM

Primitivist views of color have been offered by John Campbell, Colin McGinn, Michael Watkins, and—perhaps—by Justin Broackes.[11] But their arguments are typically negative, and can consequently leave one wondering.[12] Campbell, for example, seems primarily concerned to deflect certain criticisms of primitivism, rather than to explain in positive terms how we could possibly understand primitivism to be true. Similarly, McGinn's central thesis that colors only supervene on (without being identical to) dispositions to produce experiences of color is offered without any explanation whatsoever. Rather, his paper primarily argues that this thesis preserves what was right in his former dispositional theory without the liabilities. Yet, as Watkins rightly points out, "[s]upervenience claims do not explain, they require explanations."[13] And the bulk of Broackes's paper is devoted to showing that colors can figure autonomously in certain kinds of laws. All of these views leave it rather mysterious how there could be *sui generis* properties of the sort that colors are claimed to be, and how we could ever take our phenomenal experience to inform us about them, or about when and where they are instantiated. That is, these views can leave one equating '*sui generis*' with 'magical' and therefore as implying 'to be eliminated from our ontology'. And indeed there are some philosophers who hold that colors, as we understand them, really are *sui generis*, and that, as a consequence, we ought to be color eliminativists.[14]

We can begin to give an unmysterious primitivist account by drawing attention to the ostensive teaching of color terms, and to the role of phenomenal experience in such teaching. There can be no question but that color words—at least for the four primary colors—are taught ostensively when they are taught at all.[15] And unless one is a skeptic about other minds, it is hardly more controversial that it is essential to the correct learning of color words that one's phenomenal experience share—at least—certain structural features with that of other language learners. One might suppose that what is going on during such teaching is that color terms are given rigid physical referents, picked out by phenomenal experiences under certain circumstances, much as 'water' is given H_2O as a referent.[16] This suggestion, given the use it makes of the notion of rigid designation, presupposes that there are correct answers to questions about the colors of objects under odd counterfactual circumstances: circumstances in which light behaves differently, or in which the atmosphere of the Earth absorbs light of certain wavelengths, or in which the human visual system is different. However, given the primarily practical uses of color words, there is no particular reason to think that even the most complete understanding of their meaning would determine what to say about such cases. That is, there is no particular reason to suppose that ostensive teaching, in the case of color terms, must cause such terms to pick out, rigidly, any physical property.[17] Perhaps they do, but the above story about how such terms are taught and learnt itself provides no support for this hypothesis. And given that the hypothesis amounts to the assertion of the existence of a number of additional semantic rules governing relations between claims about color and claims about physical constitution, it seems clear where the burden of proof lies.

---

[11] Campbell, "A Simple View of Color," in Byrne and Hilbert, eds., pp. 177–90; McGinn, "Another Look at Color," this JOURNAL, xciii, 11 (November 1996): 537–53; Watkins, "Seeing Red: The Metaphysics of Colours without the Physics," *Australasian Journal of Philosophy*, lxxxiii (March 2005): 33–52; Broackes, "The Autonomy of Color."

[12] Watkins is an exception. It is his explicit project to show how colors, thought of as *sui generis* intrinsic properties of objects, might be causally responsible for experiences of color, even though they are distinct from the physical properties on which they supervene, and which also cause those same experiences.

[13] Watkins, p. 43.

[14] This seems to be the view of Paul Boghossian and David Velleman. See their "Color as a Secondary Property," *Mind*, xcviii (January 1989): 81–103, and "Physicalist Theories of Color," *Philosophical Review*, c (January 1991): 67–106.

[15] Some cultures have fewer than four basic color terms. See Brent Berlin and Paul Kay, *Basic Color Terms* (Berkeley: California UP, 1969).

If color terms *do not rigidly* designate physical properties, this should not be taken to imply that they *do nonrigidly* designate them (perhaps under the description 'physical property picked out by this kind of ostensive teaching'). Such a view, which one might regard as a functionalist account of color, would also yield determinate answers to questions that, arguably, have none.[18] It would just yield different answers than does the physicalist rigid-designation view. But if a physical surface property that might be used in our world to teach 'blue' were used in some other world to teach a word that we would unhesitatingly translate as 'yellow', there is no reason to think that there is a fact of the matter as to whether or not surfaces with that property, in that world, are yellow, or blue.[19] Whether we think of ostensive teaching as yielding a term that picks out a physical property rigidly, or nonrigidly—indeed, whether we think it picks out a physical property, a disposition, or a functional property—we will, in all cases, give color terms an intension that forces answers to questions about bizarre counterfactual cases. But, as will be argued below, there is reason to deny that a mere fixing of the physical facts is enough to underwrite the truth of any particular answer in such cases.[20]

---

[16] This seems to be Broackes's suggestion, which is why I only tentatively classify him as a color primitivist. See Broackes, p. 206.

[17] Talk of rigid designation for predicates cannot be taken too strictly on analogy with similar talk regarding names for particulars. I take the claim that a predicate rigidly designates to indicate that its extension is not determined by the descriptions that competent speakers might associate with it, but by the most general description, of the appropriate sort (physical, psychological, biological), that *actually* picks out a large enough proportion of the sample used to fix the reference of the term. For a brief and illuminating discussion of the problems of uncritically extending talk of rigidity from particulars to kinds, see Scott Soames, *Philosophical Analysis in the Twentieth Century Volume 2* (Princeton: University Press, 2003), pp. 423–56.

[18] See Cohen, "Color: A Functionalist Proposal," *Philosophical Studies*, CXIII (March 2003): 1–42.

If what has been suggested so far is correct, then what we will need is a view that makes fewer counterfactual commitments than the various forms of reductive realism. Primitivism is such a view. According to primitivism, color words do indeed rigidly designate, but they rigidly designate colors rather than physical properties, dispositions, or functional properties. Part of what one learns when one learns the meanings of color words is to apply them to objects, based on visual inspection. This implies the existence of an appropriate link between phenomenal experiences and the use of such words. But that certainly does not mean that the meaning of a word such as 'red' is 'such as to cause, in me, this particular phenomenal experience' or 'such as to cause, in normal people and normal viewing conditions, this particular phenomenal experience'. Rather, someone with a proper understanding of 'red' knows, without being able to articulate the details, that certain conditions are ones in which red objects will not produce their typical phenomenal experience. This additional knowledge, however, which can be viewed as a sort of practical ability, is by no means sufficient to determine how someone who had mastered 'red' would apply it to an object in a world stipulated to have different laws of nature, or in which human beings had slightly (or drastically) different optical equipment. That is, the meaning of 'red' does not yield answers to questions regarding the redness of objects in such worlds.

---

[19] The possibility envisioned here is one in which people classify objects exactly as we do, with the exception of objects that have the surface property of this blue (as we call it) object.

[20] Evidently Frank Jackson shares something like this view regarding the kind of reference color words carry. See Byrne and Hilbert, "Color Primitivism," *Erkenntnis*, LCVII (March 2007): 73–105, at p. 100. Mark Johnston, too, seems to hold that the meaning of color terms need not yield determinate verdicts in the relevant counterfactual cases. See his "How to Speak of the Colors," in Byrne and Hilbert, eds., pp. 137–76, at p. 155.

To some philosophers it might seem to be a liability of primitivism that it avoids commitment to the truth or falsity of the peculiar counterfactuals that often crop up in discussions of the nature of color properties. Surely if colors are properties, then there is a fact of the matter as to the color of any particular object, no matter what odd sort of world it appears in. The right response to this worry is simply to deny the assumption that grounds it. Being made of wood is a property. We are realists about wood. And yet if we imagine strange enough possible worlds we can easily come up with substances about which the best thing to say is that it is indeterminate whether or not they are made of wood. For example, we can imagine a treeless world in which a substance exactly like wood, used for houses and furniture and so on, is produced by tiny marine organisms, like coral. Is this a world in which wood is produced like coral, or a world in which there is no wood? There is no nonstipulative answer to this question. The same is true for possible worlds in which objects with the same primary qualities as red objects reflect light in a pattern characteristic of actual blue objects into the eyes of human beings whose resulting phenomenal experiences are like our experiences of yellow. What color are such objects? The point is not merely that there is disagreement among normal people, or among philosophers. Rather, it is that color concepts, like concepts of other secondary qualities but unlike those of primary qualities, have their extensions determined by the responses of human beings. When those responses fail to determine a fact of the matter, there is no fact of the matter to be determined, because there is nothing else, even in principle, to determine it.[21] And what goes for extension here goes equally for intension.

The primitivist position, despite its refusal to make certain pronouncements as to the colors of objects in odd counterfactual circumstances, does not prevent us from using color properties in our descriptions of possible worlds—even the odd possible worlds that so often figure in the internal disputes between advocates of the various forms of reductive realism. For even though 'orange' is not a rigid designator for a property that we can describe in nonchromatic terms, it is still the name of a property. Consequently, we can use it in describing a possible world in which grass is orange. What do we know about such a world? One thing we know is that the grass is yellowish and

---

[21] Similar remarks have been made, for similar reasons, regarding evaluative concepts. See Hilary Putnam, *The Collapse of the Fact/Value Distinction* (Cambridge: Harvard, 2002), pp. 108–09. See also Crispin Wright, *Truth and Objectivity* (Cambridge: Harvard, 1992); Michael Smith, *The Moral Problem* (Cambridge: Blackwell, 1994); Frank Jackson, *From Metaphysics to Ethics* (New York: Oxford, 1998).

LIBRARY UNIVERSITY OF CALCUTTA

reddish, and that it is not green Also, given plausible assumptions about the supervenience of colors on the physical (including the physical nature of human beings), we know that if the laws of nature are the same in that world as they are in ours, and if human beings have visual equipment pretty much like ours, and if the environment in which humans live is like our environment in terms of ambient light, then grass appears orange to human beings there What this suggests is that the lack of a truth value for claims about the colors of objects in odd counterfactual circumstances need not be the result of any incompleteness of meaning for color terms Rather, the lack of truth value is the result of an underspecification of the relevant world, much as Saul Kripke's "purely qualitative descriptions" of the world cannot tell us whether or not a certain person in that world is Nixon [22]

So far in this section I have tried to provide an account of the meanings of color words that fits a primitivist realist position, but not a reductive realist position My strategy has been, in essence, simply to draw attention to relatively uncontroversial facts about the teaching and learning—and hence the meaning—of color words Mastery of such words—of their meaning—primarily involves the ability to apply such words to objects in this world based on visual experience [23] It also involves the ability to apply them to objects in a limited class of nearby worlds based on regularities that hold in this world, and to distant worlds based on stipulation we can say that the meat would have been red in the center if it had been cooked less, and that if angels all have bluish eyes then none of them have yellowish ones There is no reason to suppose that the meanings of color words determine what is right to say about the colors of objects in odd counterfactual circumstances, even when we specify all the facts that *would* serve to determine the color, were any of the various forms of reductive realism true Primitivism about color allows us to ask the full range of questions of the following sort 'What if human beings all saw red things as green?', 'What if objects with such-and-so surface properties (which in the actual world are responsible for its being red) were green?', and 'What if red objects reflected light in such-and-so pattern (which in the actual world is characteristic of green things)?' Reductive realists must always regard one such question as nonsensical But given that every other reductive realist—as well as the primitivist—will give sense to that question, each denial of sense to a particular question should seem dogmatic[24] This strongly suggests that all forms of reductive realism unnecessarily precisify (and thus misrepresent) the meanings of color words

---

[22] Kripke, *Naming and Necessity* (Cambridge Harvard, 1980), pp 43–46 Kripke allows that other, *nondescriptive* features of a world might allow us to infer that a particular person is, or is not, Nixon But the parallel claim is also true for colors if we know that an object is bluish, for example, we know it is not yellowish The present point is only the uncontroversial one that the descriptions we offer of possible worlds—purely qualitative or not—can easily underdetermine various further facts about those worlds

[23] Michael Watkins's position, according to which properties are individuated by causal powers, and according to which the causal powers of colors are limited to those involving the production of color experiences, is something like the ontological image of this semantic view On Watkins's view, however, we cannot consistently talk about worlds in which colors contribute different causal powers than those they actually contribute This seems to me a liability of his position

A question that will naturally occur to the critic of primitivism at this point is 'Why doesn't all the foregoing argument apply equally well to *all* property words?' First, it should be admitted that we learn shape words partly by means of ostension There may well be approximately the same number of children's books that teach shape words by means of pictures of shapes as there are that teach color words by means of color patches Moreover, it is also true that competence with the word 'square' involves an understanding that there is a difference between 'square' and 'appears square', and therefore an implicit understanding that some conditions are such that one cannot correctly apply the word 'square' simply on the basis of the square appearance of an object In these ways an understanding of shape words is similar to an understanding of color words However, in the case of squareness there are also other necessary conditions that one learns as part of learning the meaning of 'square' For example, squares of the same size must be able to fit together in a grid Also, all the sides of a square must be the same length, and we have many methods for determining sameness of length Given the practical purposes of shape words (getting things to fit together, for example), it is not a surprise that these sorts of nonperceptual tests for squareness have a kind of priority over mere visual appearance As a result, we can construct counterfactual situations in which we have enough information to know that these additional tests would (or would not) yield a verdict of 'square' Thus, we can know, regarding these counterfactual situations, which objects are (or are not) square And this remains true even if we describe such possible situations without making any use of shape words, and even if we stipulate that there are no human beings around indeed, even if we stipulate that the only human beings around have systematic problems with visual shape-perception Interestingly, if we describe possible worlds in which our canonical tests cannot be applied, we *do* get an underdetermination analogous to that in the case of color[25] In such cases, the question of whether or not a certain four-sided object counts as square will either be a matter for stipulation, or will be as nonsensical as the question of whether a certain velocity is red

---

[24] The argument is therefore similar to one from symmetry, but with this difference In an argument from symmetry, it would matter if there were some small asymmetry that gave one position an advantage But in the present case (a) every reductive realist is outvoted by an impressive majority in at least one case, and (b) the primitivist is not outvoted in any case

Unlike shape words, color words are applied to actual objects based on nothing but chromatic experience and a practical grasp on what counts as normal conditions Even applications of color words based on memory, the testimony of others, or the output of a mechanical color-detector all "bottom out" in the verdicts of competent human judges who use nothing but their eyes to determine the colors of objects That is, a mechanical color-detector has no authority of its own it must initially be calibrated by comparison with a competent human judge Even the most hard-nosed productance physicalist has no other option in determining what set of surface productances constitutes redness Relatedly, color memory only has authority in virtue of the fact that it gets things right by the direct criterion On the other hand, we can easily imagine a machine whose output might call into question our ability to judge squareness correctly by vision alone

IV THE PROBLEM WITH REDUCTIVE COLOR REALISM

In "How to Speak of the Colors," Mark Johnston lists five core beliefs about color, and argues that no coherent view of color seems likely to preserve them all[26] In defending this claim Johnson criticizes both primary and secondary quality accounts of color In this section, I first explain how it is that primitivism avoids a tension that Johnston regards as entirely internal to the five core beliefs that is, a tension that he does not think depends on any particular realist assumptions What I hope to show is that the tension does in fact depend on a certain assumption the assumption that distinguishes nonprimitivist from primitivist views Inasmuch as the tension is both worrisome and avoidable, this argues in favor of a primitivist version of realism I then explain how Johnston's criticisms of both primary and secondary quality accounts of the colors have the same source as Benacerraf's criticism of set-theoretic definitions of the natural numbers In particular, these accounts ascribe too many properties to the target of analysis For each sort of reductive color realism, the attribution of these extra properties stems from the fact that the accounts identify the colors with something putatively more basic, just as the logicist tried to identify the natural numbers with particular sets [27] The pervasiveness of this "overascription" problem again argues in favor of primitivism

---

[25] This is part of the point of Ludwig Wittgenstein's thought experiment in which measuring or weighing things does not yield consistent results See Wittgenstein, *Philosophical Investigations* (New York Macmillan, 1953), §142

[26] The distinction between core beliefs about a subject and peripheral beliefs involves the idea that to give up a core belief is to give up the idea that one knew what one was talking about Of course it can be a matter of controversy whether a belief counts as a core belief or not See Johnston, p 137

A Tension in the Core Beliefs Here are the five core beliefs that Johnston begins with He presents them in terms of one particular color—canary yellow—and I will follow him in this

Paradigms Some of the paradigms of CY (canary yellow) are really CY

Explanation Sometimes something's being CY explains why it looks CY

Unity CY has a unique location in the space of similarity/exclusion relations for the various determinate shades

Perceptual Availability One can form justified beliefs about the CY-ness of an object on the basis of visual perception

Revelation The intrinsic nature of CY is fully revealed in an experience as of something CY

Johnston claims that there is an immediate tension among these beliefs If this is right, then the charge of failing to account for all of the beliefs will not be devastating to reductionist—or any other—views After all, no view can be expected to account for the truth of a set of inconsistent beliefs The tension Johnston is concerned with is the following Revelation cannot be true if Explanation is, given what we know about the potential explanatory causes of color experience For visual experience does not reveal the nature of these potential causes, much less which of them is the real cause [28] Visual experience does not even tell us that information about colors is conveyed through space by light It is partly this tension, and the resulting putative inability of *any* account to preserve the five core beliefs, that motivates Johnston to make his revisionary suggestion that we need to give up Revelation in the following way CY is a disposition, but an experience as of something CY does not reveal this fact about its intrinsic nature.

---

[27] Boghossian and Velleman offer similar arguments against both sorts of reductive accounts See their "Colour as a Secondary Quality," and "Physicalist Theories of Color," both reprinted in Byrne and Hilbert, eds , pp 81–103, and pp 105–36 Rather than opting for Johnston's revisionary dispositionalism, Boghossian and Velleman advocate an error theory according to which we mistakenly attribute to objects properties that belong only to visual experience

[28] Johnston, pp 139–40

Johnston's arguments make use of a relatively strong version of Revelation: a stronger version than we need accept. Moreover, since Johnston's version of Revelation conflicts obviously and immediately with reductionist accounts of all sorts, it is to be expected that advocates of such accounts will reject it. Further, Johnston's Revelation is a controversial thesis: indeed, it is somewhat controversial what precisely it amounts to. For all of these reasons, I want to present a weaker version of Revelation that I think is true. This weaker version will ultimately yield the same anti-reductionist implications as the stronger version. So perhaps reductionists will simply reject it. But since it is weaker, it should be harder to reject. My version is also sufficiently strong that Johnston will still advocate a view that violates it. That primitivism need not violate it is therefore something in its favor. The version of Revelation I endorse is the following:

> *Modest Revelation*: A complete understanding of what CY is can be given by an experience as of something CY, in someone with normal human vision who has already acquired color concepts.[29]

This view implies and is implied by what seems to be true: that if someone has normal human vision, and knows how to use color vocabulary, then we can explain what CY is to that person completely simply by showing her something that is CY and saying that the color of the object is CY. This simple instruction is all that is needed to give her complete competence with the term 'CY'. Moreover, the same capacities that underlie the efficacy of this simple method of instruction would also allow our subject to coin her own term when observing a CY object a term that meant the same as 'CY'.

---

[29] Modest Revelation does not make use of the phrase 'intrinsic nature', since it seems that intrinsic nature is not what Revelation is about, at least if intrinsic nature and extrinsic nature are meant to be exclusive and exhaustive. For something (a brick) might, as a matter of fact, be part of the intrinsic nature of something else (a house), and yet be only contingently so. I think proponents of Revelation are in fact concerned with *essential* nature, which cannot be contingent, and which I indicate with the less technical phrase 'what CY is'. It may be useful to compare Modest Revelation with the following passage from Wittgenstein:

> the ostensive definition explains the use—the meaning—of the word when the overall role of the word in the language is clear. Thus if I know that someone means to explain a color-word to me the ostensive definition 'That is called "sepia"' will help me to understand the word (*Philosophical Investigations*, §30).

The similarity suggests Modest Revelation will not fall foul of Wittgensteinian considerations having to do with the nature of concept possession and its normative aspects. This is a danger that stronger versions of Revelation seem to ignore. Indeed, standard versions of Revelation seem to suggest that babies, apes, and dogs have full knowledge of the nature of the colors they see. But, to borrow again from Wittgenstein, "looking does not teach us anything about the concepts of colours" (*Remarks on Colour*, G.E.M. Anscombe, ed., L. McAlister and M. Schättle, trans. (Berkeley: California UP, 1978), Part I, §72).

Modest Revelation is considerably weaker than Revelation. But for that reason it is much more plausible. One way it is weaker is that its scope is restricted to people who have already acquired color concepts and who can therefore be understood to be "paying attention to the color" of an object when it is shown to them. This makes it far more plausible that they can learn about the nature of a particular color via an experience of that color. Advocates of the stronger version of Revelation seem to think that we are simply born with the notion of a color, and only want acquaintance with colored objects to know *which* colors there are.[30] The further restriction to people with normal vision is warranted for the following reason. Consider Bob, who is red/green color blind. Bob, looking at a yellow object that does not seem reddish or greenish even to someone with normal vision, may well have an experience of that object that is phenomenally indistinguishable from the normal experience.[31] And yet, given that Bob cannot reliably distinguish yellow objects from many nonyellow ones, it seems wrong to say that the experience he has when looking at the unique yellow object reveals to him what unique yellowness is.

Certainly Modest Revelation is in tension with the claim that CY is a disjunctive physical surface property, or a set of surface spectral reflectances. In general, it is in tension with the claim that CY is any property that, as a matter of fact, causes, via the propagation of light, experiences of CY. But this point only makes Modest Revelation inconsistent with Explanation if we hold that the only account of the way in which colors explain color experiences must be given in terms of the propagation of light, the absorption of photons by photopigments, electrochemical interactions between neurons, and so on. But all that Explanation actually claims—and all that it should claim, if it is to count as a core belief—is that sometimes something's being CY explains why it looks CY. There are many forms of explanation other than those that appeal to the basic laws of physics Thus, knowing what CY is does not entail, even if Explanation is true, that one knows that CY is a microphysical surface property There are many forms of explanation that have a legitimate claim to be called causal explanations perhaps psychological explanations count as causal, and appeals to the gross shapes of objects certainly seem to provide causal explanations for the behavior of those objects without any appeal to microphysics What matters in all such cases is the existence of the right sort of counterfactual-supporting law-like generalizations Primitivism can take advantage of this fact to avoid the tension Johnston describes For on a primitivist account of color, it remains true that the colors of objects explain, at least quite often, why those objects appear to have the colors they do [32] Admittedly, one's current visual experience may in principle be explained by reference to the propagation of light, the functioning of one's eyes and optic nerves, and one's brain No mention of color need enter into such an explanation Let us grant this But in fact the relevant question is not 'What is the explanation of my current visual experience?' Rather, it is 'What explains why I am having an experience of CY?' True, my current experience happens to be one of CY And, we are assuming, we can explain how my current experience is caused without reference to the property of being CY But this does not mean that we have an explanation of why I am currently having an experience in which something looks CY [33] Explanatory contexts, like epistemic contexts, are opaque Assume for the moment that the event that consists in my saying 'No' in response to some proposal can be explained in terms of physical processes originating in the utterance of the proposal and terminating with the contracting of certain muscles in my mouth, lips, and tongue Such an explanation would not address the question of why I rejected the proposal, even if (let us grant) the event *was* my rejection of the proposal A purely physical explanation of my current experience (as I look at a CY object in good light) is simply not an explanation of why I am having an experience *of CY*

---

[30] My very small friend Ben Carens seems to me to have falsified this supposition. When he was first learning color words and had only got as far as red, yellow, blue, and green, I showed him an orange object and asked its color. His answer: "No color."

[31] As many have suggested, it is problematic to talk about interpersonal indistinguishability of phenomenal states. But those who endorse a strong version of Revelation seem to assume that it is "the particular phenomenal quality" given in an experience of CY that reveals the nature of CY. So, in arguing that their version of Revelation is too strong, it is legitimate to make use of the assumption that we can talk about particular phenomenal qualities in this way.

---

[32] See Campbell, pp 182–84, and Broackes, pp 192–94, for more detailed development of this same point

[33] Gabriele De Anna, in defending the primitivist position, ignores this point in a revealing way He offers the critic of primitivism a sort of translation of the claim that the red of an apple caused the red of a patch on a photo a translation that does not mention colors at all But in fact De Anna's translation does not really manage to omit reference to colors rather, it conceals it a footnote that contains the following text "[t]he relevant similarity here is one which allows both s' and s'' to be considered experiences of red " See De Anna, "The Simple View of Colours and the Reference of Perceptual Terms," *Philosophy*, LXXVII (January 2002) 87–108, at p 99

Suppose we bracket the above concern, and grant that we could have a physical explanation of my experience of CY To see why such an explanation would still not be as good as the primitivist explanation, consider the following question If the surface had been physically different, even radically so, would this entail that my experience would not have been an experience of CY? No Very different surface properties—indeed, very different reflectance properties—can easily give rise to the same phenomenal experience On the other hand, had the object not been CY, it is unlikely I would have had an experience of CY This suggests that it is the property of being CY that will figure in the kinds of regularities characteristic of causal explanations [34] So primitivism allows us to claim both that the color of an object often explains why it looks the way it looks, and that the nature of particular colors can be fully revealed, to those competent with color concepts, in experiences of those colors Nonprimitivist realist views cannot underwrite both of these claims This makes the charge of failure to account for the five beliefs more damaging to nonprimitivist views, and argues in favor of primitivism

Primary Quality Accounts Too Many Dimensions of Similarity Let us now turn to Johnston's criticism of primary-quality accounts of color The argument proceeds in two stages [35] First, Unity requires that CY have its unique location in color similarity space it should be more like other shades of yellow than it is like any shade of blue, and so on But on the primary quality account, these relations are going to be a matter of scientific discovery about similarities between massively disjunctive primary qualities This, however, is inconsistent with the fact that we knew all about color similarities before science told us anything of this sort So colors cannot be primary qualities [36] Thus ends the first stage of the argument The second stage starts from a potential reply to the conclusion of the first stage Might it not be, according to the reply, that the familiar relations—the ones we knew before science told us anything about the colors—function as *a priori* conditions that will constrain the properties that science could discover to be the colors? That is, science must look for properties such that CY is more similar, in terms of those properties, to other shades of yellow than it is to any shade of blue But Johnston convincingly argues that with this stipulation we give up the idea that we know if there are really any CY objects at all For science would need to show that there really is a primary quality that bears the right sort of similarity relations to other colors And, for all we know, there is no such primary quality The real possibility that there might not turn out to be any CY objects violates Availability Thus, the primary quality account—adjusted to meet Unity—violates Availability

---

[34] Byrne and Hilbert make these and related points on behalf of the primitivist in their paper criticizing primitivism See Byrne and Hilbert, "Color Primitivism " Watkins also provides a very nice explanation of the same point in "Seeing Red," pp 47–48

[35] Johnston, pp 149–50

[36] See also Hardin, p 66

One might try to avoid this problem by claiming that there could be more than one conceptually independent but noncontingent description of the colors one in terms of location in similarity space, and another in terms of primary qualities A model from another domain might be 'H_2O' and 'the actual watery stuff' [37] That would allow us to know that a certain object was CY, simply by knowing about its visual similarity relations, even if we did not know anything about its microphysical surface properties, and even if no such properties had the same similarity relations The problem with this proposal is that Unity is meant—plausibly—to be a necessary truth For it does not seem possible for orange to be visually more similar to blue than to either yellow or red But if orange is supposed to be a certain microphysical property, we are forced to give up the idea that something orange must appear more similar to something red than to something blue

The conflict between Unity and Availability might not seem to be the result of the attribution of too many properties to the colors But in fact that is one way of viewing the source of the problem If colors are actually microphysical properties—disjunctive and messy or simple and unified—then they will stand in various microphysical similarity relations to which we have no access except via quite advanced scientific techniques That is, on the physicalist account it might turn out that a certain shade of blue is actually more similar to a certain shade of yellow than to any other shade of blue This is because the physicalist gives colors many more dimensions of similarity to other colors than any normal person would typically be willing to grant For the only similarities that normal people attribute to colors are visual similarities similarities in hue, saturation, brightness, and so on The extra dimensions of similarity that the physicalist imports are like the extra relations that the logicist mistakenly gave to the natural numbers membership relations to each other The proper response is to deny that colors have any microphysical similarity relations And this is to deny that colors are in fact to be identified with microphysical properties

---

[37] I owe this objection to Cohen

Dispositional Accounts Too Many Relata Johnston next criticizes dispositional accounts of the colors Here it is Revelation that does the bulk of the work For the relevant dispositions are relational properties they are dispositions to cause certain experiences in a certain sort of being in certain circumstances But, Johnston argues, most colors do not appear to be relational There are perhaps certain exceptions The highlights on a shiny object may change and shift as one moves in relation to the light source, and this may make the relativity to circumstances phenomenally apparent But many colored objects simply appear to have a steady color that does not vary except with very drastic changes in the light Again, this failure of an account of color to meet one of Johnston's five core beliefs might not immediately seem to be the result of the fact that the account attributes too many properties to the colors But, again, the source of the problem is exactly this For the failure of Revelation for *any* account of color is typically explained in the following way on that particular account of color, the color of a particular object has certain essential properties, but those properties are not fully revealed by the chromatic aspects of an experience of that object In the case of dispositional accounts, the color of a particular object has a number of relata that are not revealed in experience The general moral here is that any failure of an account in terms of Revelation—or Modest Revelation— can often naturally be put in terms of that account's ascribing too many properties or relations to the colors Indeed, if an account of the colors makes it essential to a particular color that it have *any* properties or relations that an experience of that color would not reveal to those competent with color concepts, that account will be inconsistent with Modest Revelation So Johnston's criticisms, to the degree that they can be made to rely on Modest Revelation, can easily be understood as a general indictment of all nonprimitivist realistic accounts of color That is, it can be understood on the model of a general indictment of set-theoretic reductions of the natural numbers that go as far as to claim that the numbers are nothing but the sets that the account identifies with them Just as the Fregean account allowed for answers to such ill-posed questions as 'How many members does five have?', so too does the dispositional account of color allow for an answer to the ill-posed question 'Relative to what viewers does that object have the color it appears to you to have?'[38] This question is ill-posed because no class of viewers appears, even obliquely, in the content of chromatic experience.[39] This objection might also form the basis of an objection to relationalist views such as those of Jonathan Cohen and Brian McLaughlin. On their view, claims such as 'That apple is green' are much like the claim 'That refrigerator is too heavy to lift'. The latter claim conceals a reference to some individual or class for whom the refrigerator is too heavy, and Cohen and McLaughlin hold that a similar reference is part of the content of the former claim.[40] By itself this view does not offend against Revelation. But if Cohen and McLaughlin think that such common color claims serve to express even part of the content of visual experience—if they think, that is, that part of the content of a visual experience might be 'there is a red object near a green object'—then they are committed to a view that is incompatible with Modest Revelation.

---

[38] This question is quite different from the following 'What viewers have the same phenomenal experience as you do, when looking at the object?' For even if two viewers have the same phenomenal experience, the dispositional theory cannot say that those viewers see them as having the same color The reason for this is that, even assuming

All Reductive Realist Accounts: Too Many Modal Properties. In a recent argument against primitivism, reflectance physicalists Alex Byrne and David Hilbert begin by assuming that the primitivist must choose one of two ways to explain the connection between colors and reflectances.[41] The primitivist must say, according to Byrne and Hilbert, either that colors are nomologically coextensive with reflectance types, or that, for any color, there is some reflectance type that metaphysically necessitates that color. One essential difference between these two explanations is that the former makes the connection contingent, since laws can change from world to world, while the latter makes the connection necessary. Colin McGinn is a representative primitivist who embraces the nomological understanding of the connection. According to McGinn, if the laws of nature in some other possible world result in its being the case that ripe tomatoes

---

that the relevant viewing conditions *C* are identical in 'CY-for-humans-in-*C*' and 'CY-for-Martians-in-*C*', these two dispositional properties clearly are not identical, and need not be coextensive. Thus, even if a certain object *O* has both of these properties, the fact that you and a Martian have the same phenomenal experience when looking at *O* does not mean that, relative to Martians, *O* has the color property it appears to you to have. Compare John Hyman, "What, if Anything, Are Colours Relative to?" *Philosophy*, LXXX (October 2005): 475–94, at p. 486.

[39] Some sorts of relativity might be part of the content of visual experience: for example, the blueness of a certain object might appear bright relative to that of another object, but dim relative to that of a third. But this sort of relativity does not speak against Modest Revelation, since it is actually manifested in visual experience.

[40] McLaughlin, "The Place of Colour in Nature," in D. Heyer and R. Mausfeld, eds., *Colour Perception: Mind and the Physical World* (New York: Oxford, 2003), pp. 475–502; Cohen, "Color Properties and Color Ascriptions"; Sydney Shoemaker, "Phenomenal Character," in Byrne and Hilbert, eds., pp. 227–45, at p. 234.

[41] Byrne and Hilbert, "Color Primitivism," pp. 75–76.

produce color experiences quite different from the experiences they produce in this world, then in that world ripe tomatoes would not be red, even if they reflected light in precisely the same way. I do not know of any primitivist who explicitly advocates the other, metaphysical understanding of the connection between colors and reflectances. But such an advocate would have to claim that the tomatoes in the previous example, if they really reflected light in the same way, would remain red no matter how they appeared to people in some other possible world.

We should note, before going on to discuss the relative merits of these two options for the primitivist, that it is not only primitivists who have a choice here. Any reductive realist will identify colors with some features of colored surfaces. Those features may simply be reflectance types, in which case the metaphysical version of the explanation of the connection between colors and reflectances will be true. Or the features may be the highly disjunctive physical properties of surfaces that give rise to facts about reflectances, in which case a red object in another possible world might have a different reflectance, since the laws of the propagation of light might be different in that world. In that case the nomological version of the connection between colors and reflectances will be true. All other reductive views have similar implications. As a result, reductive realists are in a more tightly constrained position than primitivists here. They cannot avoid endorsing one of the two understandings of the connection between reflectances and colors, because their views as to the nature of color will force that endorsement. This is because all the distinct reduction bases for the colors that are not metaphysically coextensive are nomologically coextensive. To make this clearer, imagine that the various potential reduction bases for redness are *A*-ness, *B*-ness, *C*-ness, and so on, and that none of these reduction bases is metaphysically coextensive with any other: their coextensiveness is nomological. If redness simply is *A*-ness, then redness and *A*-ness are metaphysically coextensive because they are identical, while redness and *B*-ness, *C*-ness, and so on, are nomologically coextensive, because *A*-ness (which is redness) is nomologically coextensive with all the other potential reduction bases. All plausible claims of the form 'Redness is coextensive with *X*-ness' will, therefore, be either nomologically or metaphysically necessary.

There is no need here to discuss the details of the argument that Byrne and Hilbert mount against the primitivist. For the central point of this section is that we can reject the very first premise of that argument: primitivists, contrary to what Byrne and Hilbert assume, need not make any choice between the two explanations offered to them. That is, the primitivist need say neither that redness is metaphysically necessarily coextensive with a particular reflectance type, nor that redness is nomologically coextensive with that reflectance type. By this I mean the strong claim that the primitivist can simultaneously deny both explanations of coextensiveness—not merely that the primitivist can withhold judgment until the facts are in, or deny either as long as he endorses the other. This does not mean that the primitivist must deny both explanations. Rather, either choice remains open to the primitivist, so that if an argument should emerge that decisively favors the metaphysical account, for example, over the nomological account, the primitivist can simply accept it.

How is it that the primitivist can deny both explanations? Surely the connection between colors and reflectances is either a necessary connection, or is contingent on the laws of nature! What third option could there be? The answer is that commitment to one understanding of the connection or the other is tantamount to the endorsement of certain counterfactuals. An advocate of the nomological explanation of coextensiveness, such as McGinn, will hold that it is because of the relevant physical (and perhaps psychophysical) laws that objects with certain reflectances cause certain experiences in normal human beings, and that facts about the colors of objects supervene on the resulting dispositions of objects to produce experiences. Thus the following counterfactual is true:

> *N* If the relevant laws were such that an object with reflectance in class *X* (characteristic of actual red objects) produced experiences in normal human beings that were quite different from the experiences it actually produces, then that object would have *a different* color from that which it has in the actual world.

On the other hand, the advocate of the metaphysical understanding of the connection would deny this. That is, on the metaphysical understanding, the following counterfactual is true:

> *M* If the relevant laws were such that an object with reflectance in class *X* (characteristic of actual red objects) produced experiences in normal human beings that were quite different from the experiences it actually produces, then that object would have *the same* color as that which it has in the actual world.

Note that although the consequent of *M* is simply the denial of the consequent of *N*, this only implies that *N* and *M* exhaust our options on the assumption that any given counterfactual must be either true or false. This assumption is clearly false, since—to put it in Lewisian terms—it implies that there is always a unique closest antecedent world. Thus, in order to deny both explanations, the primitivist need only deny a determinate truth value to each of these counterfactuals.

This is a live option that is not available to reductive realists.[42] Moreover, it seems to be the correct option, since it is the only option that allows us to talk both about worlds in which red objects appear blue (falsifying *N*), and about worlds in which red objects reflect light differently (falsifying *M*).

Many realist positions in other areas allow us to deny truth values to all of the members of analogous sets of seemingly exclusive and exhaustive counterfactual claims (typically either/or pairs). This is often the result of the fact that the meanings of the relevant property terms, in combination with what is specified in the antecedents of the counterfactual claims, underdetermines truth value. One explanation for this underdetermination has already been mentioned: the antecedent world may simply be underspecified. This is what is going on in Kripke's Nixon cases. A second explanation is that it is not necessary, for many property terms, to have ready rules determining their use in situations that never actually arise. For example, in our world people never split into two separate conscious beings. Nor do two separate people ever fuse into one person. As a result there is no need for rules governing the identification of people through such transformations. And, as a further result, there are no such rules, and no determinate truth value for claims about the identity of persons in counterfactual cases involving human fusion and fission. These truth-value gaps make absolutely no difference to our everyday use of the concept of a person, and are perfectly compatible with the literal truth of the claim that the author of this paper, and its readers, are people. It is quite implausible to say that our concept of a person is therefore, and "strictly speaking," incomplete.

V. OBJECTIONS

Primitivism is certainly still a minority view, despite its ability to avoid the kinds of errors that are characteristic of reductionist views: errors that either stem from or result in the attribution of too many properties to the colors. Because of its minority status, it is to be expected that it will meet with resistance. In this final section I want to deal with two objections to primitivist views that were not dealt with in passing above.

1. Spectral Inversion and Indeterminacy of Referent. David Chalmers offers an objection to color primitivism that involves the possibility of an inverted spectrum[43] Putting it in Chalmers's terms for the moment, the objection involves the possibility of a person, or group of people, whose everyday color experience is inverted in relation to ours These people, looking at grass, have experiences we would describe as of something red, and, looking at blood, they have experiences we would describe as of something green Of course there are host of objections to this way of describing matters If we assume that the inversion is supposed to be behaviorally undetectable, then there are powerful arguments that the inversion described is completely irrelevant to any claims about color at all That is, on the assumption of undetectability, there is a strong case to be made that these people mean precisely what we mean by 'red', 'green', and so on, and that what they have, and what we have, when we look at a red object is (typically) appropriately described as 'an experience of red'[44] But let us leave this aside for the moment, and make, with Chalmers, two assumptions

---

[42] Mark Leon comes to his view largely as a result of ignoring this possibility. See his "Colour Wars: Dividing the Spoils," *Philosophy*, LXXVII (April 2002): 175–92, at pp. 182–83. Leon is committed to the view that we would say that the colors of objects had changed, even if all that had happened was a systematic change in our own visual systems.

(1) In the actual world, people who are not color-blind represent the colors of objects correctly in most of the cases in which we would say that the viewing situation is normal
(2) Because the inverted people have different qualitative experiences when looking at grass and blood, the color primitivist must hold that they are attributing different color properties to those things

Given these two assumptions, it is natural to ask, as Chalmers does,

> when a subject in the [inverted] community has a phenomenally red experience of the apple, and a subject in [our] community has a phenomenally green experience of the apple, which of these experiences is veridical?[45]

Given Chalmers's assumptions, it is reasonable to be led to his answer 'neither' But there is no reason to accept Chalmers's second assumption According to some plausible views, and as long as the inversion is behaviorally undetectable, there is no problem with the answer 'both' Chalmers rejects this answer because he supposes that it implies that grass would, in such a case, be both red and green a situation which I am willing to grant is impossible But it is possible to answer 'both', not because one holds the apple to be both red and green, but because one holds that both communities attribute the very same color property to the apple. This is easily seen by taking Wittgenstein's private language argument just seriously enough to realize that the particular qualitative experiences of human beings play no role at all in the teaching or learning of the meanings of color words.[46] If one half of the human population were, unbeknownst to us, actually undetectably inverted with respect to the other, that would nevertheless be no reason to suppose that, also unbeknownst to us, the word 'red' really had two meanings. Indeed, it seems to me that we have no reason whatsoever to be confident that we have any insight at all into the intrinsic qualitative experiences that other people have (unless such phrases as 'an experience in which something appears green'—which have a public use—counts as a description of an intrinsic qualitative experience), and that this lack of insight is of absolutely no importance to the question of whether or not we all mean the same thing by 'red' (or by 'an experience in which something appears red').

---

[43] See Chalmers, "Perception and the Fall from Eden," in Tamar Gendler and John Hawthorne, eds , *Perceptual Experience* (New York Oxford, 2006), pp 49–125 A similar objection is raised by Jim Edwards in "The Simple Theory of Colour and the Transparency of Sense Experience," in Crispin Wright, Barry Smith, and Cynthia Macdonald, eds , *Knowing Our Own Minds* (New York Oxford, 1998), pp 371–89

[44] See Galen Strawson, "Red and 'Red'," *Synthese*, LXXVIII (February 1989) 193–232

[45] Chalmers, "Perception and the Fall from Eden," §6

Of course, we have plenty of reason to believe that certain things—cherries and raspberries, for example—look red to other people: we have no reason to doubt their reports, for example, or their linguistic competence with the relevant phrases. Nothing in the Wittgensteinian position challenges the view that most people perceive colors correctly most of the time, in the sense that red things typically look red to them, and yellow things yellow. Rather, the Wittgensteinian position holds that we have no reason to think we have insight into what it is like for other people to have experiences in which objects look red—except of course that they are experiences in which objects look similar, in color, to red things. One might try to appeal to similarities in the eye and brain to support the view that we *do* have reason to believe that other people have intrinsic qualitative experiences that are very similar to ours.[47] In such a case one would be tempted to deny the existence of variation in qualitative experience within the actual human community unless genes were discovered that resulted in distinct connections in the retina, or some other variation in the neural pathways that lead from the retina to the visual centers of the brain. But we understand so little of the connection between brain function and phenomenal consciousness that we have no reason to deny the possibility that genetics leaves the final step from neurology to phenomenology completely undetermined. After all, as long as an organism makes the correct color distinctions, there does not seem to be any evolutionary pressure to have any given sort of phenomenal experience when looking at something red. I grant that strict inversions, in which all the same hues are represented in the visual space of the inverted person, might not be possible. The reason for this is that human color space is quite asymmetrical: an inverted person of this sort would actually make different color classifications.[48] But if such inversions are ruled out, Chalmers cannot use them either, and must make use of thought experiments in which people have completely alien color spaces, or color spaces that share only a subset of hues with our own color space (or with *one's* own color space, if the reader is skeptical about our insight into the linguistically inaccessible phenomenal experiences of other people).

---

[46] See Wittgenstein, *Philosophical Investigations,* §243–93. See also Strawson, "Red and 'Red'."

[47] Byrne and Hilbert seem tempted to argue in this way. See their "Color Primitivism," pp. 88–89.

One might try to modify Chalmers's objection to make use of an inversion that is not undetectable. But if we imagine a group of people who classify objects differently than we do, then there is little reason to worry over Chalmers's rhetorical question. We should simply say that these other people see different colors than we do, just as cats and giraffes do.[49] The question as to which species perceives color correctly is obviously ill conceived if it is posed in a way that excludes the answer 'They all do'. In a similar way, in this modified version of Chalmers's thought experiment, so too is the question of which subgroup of human beings do.

2. *Animal Colors.* Alex Byrne and David Hilbert argue that primitivists cannot provide an adequate account of the discriminatory capacities of animals to which we would ordinarily attribute color vision. Goldfish, to borrow an example from Byrne and Hilbert, share many of the mechanisms of human color vision, but can see some way into the near ultraviolet, where human vision cannot penetrate. According to Byrne and Hilbert, there is a tension between primitivism and the extremely plausible idea that goldfish are responding to colors. The source of the tension stems from primitivism's connection with the thesis of Revelation. Revelation tells us that we know everything essential about the particular colors we see, just by seeing them. But if goldfish are responding to colors, and making color distinctions, it is extremely plausible that they are not making

---

[48] For a nice account of the problems involved in imagining such an undetectable spectral inversion, see Hardin, pp. 139–42.

[49] For more argument to this effect, see Watkins, "Do Animals See Colors? An Anthropocentrist's Guide to Animals, the Color Blind, and Far Away Places," *Philosophical Studies*, xciv (June 1999): 189–209.

those distinctions by responding to the colors we humans see. What colors are they responding to, then? Byrne and Hilbert think the primitivist has no good answer to this question, because the primitivist must say that the only properties that share a common nature with human colors—which goldfish colors would have to do, if they were to count as colors—are in fact the human colors. This is because, according to Byrne and Hilbert, "careful reflection on (human) color experience seems, if anything, to *exclude* the possibility of colors that are not located within the familiar color solid."[50] The only way for primitivists to claim that goldfish are responding to colors, according to Byrne and Hilbert, is to claim that they are in fact responding to *human* colors. But in that case, they are responding in a wildly bizarre way! And the only way to avoid this unpalatable result, according to Byrne and Hilbert, is to deny that goldfish are responding to colors at all. Rather, the primitivist could say that while humans respond to colors, goldfish and other animals are responding to other properties. Reductionists, on the other hand, who identify colors with, for example, sets of spectral reflectances, can easily accommodate goldfish colors. For they can simply say that any set of spectral reflectances counts as a color, and those that goldfish respond to in a systematic way are goldfish colors.

In making this criticism of primitivism, Byrne and Hilbert seem to be confusing our capacity to represent novel colors to ourselves (which we cannot do) with our capacity to imagine that there are colors we cannot see (which we can). Thus the primitivist should simply deny Byrne and Hilbert's undefended—and independently implausible—assertion that reflection on our color experience seems to exclude the possibility of colors outside of human color space. If a primitivist endorses Modest Revelation this need only commit her to the view that those of us who see orange things are aware of the essential nature of orange. One essential feature of orange is that it is a uniform property of opaque surfaces and transparent volumes. Goldfish colors will share this essential feature. What goldfish colors look like however, we will never know.

JOSHUA GERT

Florida State University

---

[50] Byrne and Hilbert, "Color Primitivism," p. 94.

COMMENTS AND CRITICISM

RISKY BUSINESS EVOLUTIONARY THEORY AND HUMAN ATTITUDES TOWARD RISK—A REPLY TO OKASHA*

In a recent article, Samir Okasha has given a novel and interesting explanation for the pervasiveness of risk-aversion in decision making—one based on evolutionary considerations [1] This explanation furthermore has considerable intrinsic importance, as it shows how useful evolutionary approaches can be for the investigation of many long-standing issues in philosophy and in the (natural and social) sciences

However, despite these successes, a closer look at the problem Okasha is trying to solve makes clear that, overall, he has not managed to account adequately for all the issues that need to be addressed As this paper aims to show, by disregarding the fact that *risk-loving* behavior, too, is widespread, Okasha's account ultimately fails to explain the nature of our attitudes toward risk Further, while a nonadaptationist extension of his account may be available, it requires much further work to be made plausible

The paper is structured as follows Section I briefly revisits the argument of Okasha's paper Section II raises an objection to Okasha's account and briefly suggests a possible reply to this objection I conclude in section III

I OKASHA'S ARGUMENT FOR THE EVOLUTION OF RISK-AVERSION

The problem that Okasha takes to require a solution is the following It is an empirical fact that people frequently behave in ways that suggest *risk-aversion* they are not only concerned with the *expected value* of the outcomes of the options under consideration, but also consider the *variance* associated with these outcomes [2] What requires explanation is the origin of this behaviour–why are people often willing to forgo potential, but uncertain, gains?[3]

---

* Many thanks to Elliott Sober, Samir Okasha, and the editors of this JOURNAL for useful comments on previous drafts of this paper

[1] Okasha, "Rational Choice, Risk Aversion, and Evolution," this JOURNAL, CIV, 5 (May 2007) 217–35

[2] In fact, disregarding the variance is also theoretically dubious (as is shown, for example, by the St Petersburg Paradox)

[3] See Okasha, pp 217–18 Note that Okasha frames this as an issue about the plausibility of Rational Choice Theory (which assumes *risk-neutrality* and thus clashes

0022-362X/08/0503/156–65

In order to answer this question, it is best to begin by making explicit what it means to exemplify a specific attitude toward risk[4] Consider a lottery L and a random variable X that ranges over the possible outcomes of that lottery Then

> The agent is *risk-neutral* if and only if she is indifferent between receiving the expected payout of L with certainty and receiving L itself, that is $E[U(X)] = U(E[X])$
>
> The agent is *risk-averse* if and only if she (strictly) prefers receiving the expected payout of L with certainty to receiving L itself, that is $E[U(X)] < U(E[X])$
>
> The agent is *risk-loving* if and only if she (strictly) prefers receiving L itself to receiving the expected payout of L with certainty, that is $E[U(X)] > U(E[X])$

Furthermore, distinguish three types of explanations of our attitudes toward risk *purely psychological* ones (which appeal only to facts about our minds now), *social scientific* ones (which give particular prominence to the constraints put on us by our social environment), and *evolutionary* ones (which appeal to the circumstances under which our minds evolved)[5] Since Okasha's interest is mainly in the last category, he groups the first two types of explanations together, and since nothing much hangs on this distinction in the present context either, I shall do the same[6]

As Okasha further notes, a pivotal position in debates about the nature of our attitudes toward risk is held by a psychological / social scientific account that is based on the assumption that people have *declining marginal utility of money* Making this assumption entails that the associated utility function is *concave*, and that weighted averages

---

with the empirical findings of risk aversion) However, since this raises difficult questions about the interpretation of Rational Choice Theory that do not bear on the actual argument at stake, it is better to frame the problem as a purely explanatory one, as is done here

[4] In what follows below, it is easiest to understand "payoff" as money However, other interpretations are possible, as long as the payouts are not directly in utility terms (the last would trivialize the issue) For similar definitions of the attitudes toward risk, see Andreu Mas-Colell, Michael Whinston, and Jerry Green, *Microeconomic Theory* (New York Oxford, 1995)

[5] It is not clear that these approaches necessarily need to conflict, but it is helpful for now to follow Okasha in assuming that they do Section III discusses the reasonableness of this assumption further

[6] For an example of a social scientific explanation of risk-aversion, see Olivier Mahul, "The Output Decision of a Risk-Neutral Producer under Risk of Liquidation," *American Journal of Agricultural Economics*, LXXXII, 1 (2000) 49–58

between two sums of money are associated with higher levels of utility than weighted averages of the utilities associated with the two sums, that is, $E[U(X)] < U(E[X])$ Graphically, this can expressed as follows

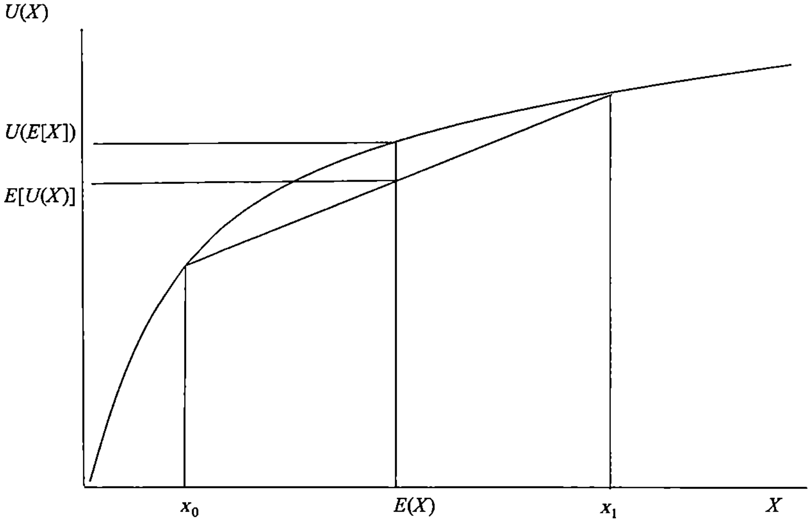


Figure 1

However, as Okasha also makes clear, this type of approach is far from being universally accepted—in particular, it has been criticized for appearing to make numerous empirically false predictions[7] For this reason, he is led to pursue an alternative explanation of risk-aversion based on *evolutionary considerations*

Specifically, he argues that, if the populations that hominids evolved in were finite or contained reproductive dependencies (so that the reproductive success of an organism in a population is correlated with that of the others), risk-averse behavior can be fitter than risk-neutral behavior[8] To make this argument, Okasha appeals to the fact that *fitness* is not always captured well by the expected ***absolute*** number of offspring an organism has Instead, it at times needs to be seen as its expected *relative number* of offspring[9]

---

[7] See, for example, Matthew Rabin, "Risk-Aversion and Expected Utility Theory," *Econometrica*, LXVIII, 5 (2000) 1281–92, and Rabin and Richard H Thaler, "Risk Aversion," *The Journal of Economic Perspectives*, XV, 1 (2001) 219–32 It is, however, worthwhile to note that Rabin's work is still somewhat controversial and that some of the assumptions he makes seem fairly strong In particular, he assumes that utility is concave *throughout*, which, as will become clear below, does not appear altogether plausible For further discussion, see also Martin M Monti, Simon Grant, and Daniel N Osherson, "A Note on Concave Utility Functions," *Mind and Society*, IV (2005) 85–96

This can be illustrated with a case of the following sort assume there are two types of organisms, A and B Organisms of type A consistently double their numbers in every new generation, while organisms of type B either quadruple (with probability 1/2) or fail to reproduce at all (also with probability 1/2) Finally, assume that A and B start out being equally numerous, and that there are n individuals of each type Then it holds that

Expected *number* of offspring of type $B = 1/2\ (4n) + 1/2\ (0n) = 2n =$ Expected *number* of offspring of type A

The expected *ratio* of A's to B's $= 1/2\ (n/6n) + 1/2\ (n/n) = 1/12 + 1/2 = 7/12 > 1/2$

What this shows is that, over time, organisms of type A are expected to increase in frequency *relative to* the number of organisms of type B, since the latter's chances of reproduction are more variable than the former's This therefore makes clear that, from an evolutionary point of view, there is selection *against* such variability in reproductive success In this manner, Okasha also comes to endorse John Gillespie's measure of fitness, according to which it holds that (for an organism of type X and total population size N)

Fitness of organism type $X \approx E(\text{number of offspring of type } X) - 2[\text{Var}(\text{number of offspring of type } X)/N]$[10]

For what follows below, it is important to be clear about what this equation expresses according to the Gillespie/Okasha measure, the fitness of an organism is its expected number of offspring, *corrected for* the riskiness of obtaining this outcome That is, an organism's fitness is *lower* the *greater* the variance of the strategies involved increases in the variance of the strategy *always* impact the fitness of the organism *negatively*.

---

[8] In what follows, I shall largely ignore these two assumptions, as they do not play any role in my argument For infinite populations, the benefits that risk-averse populations have over risk-neutral ones go to zero, but they never become *negative* The same holds if it is assumed that individuals reproduce independently from one another See also Okasha, pp 225–26

[9] For more on the concept of fitness in evolutionary biology, see Elliott Sober, "The Two Faces of Fitness," in Rama Singh, Diane Paul, Costas Krimbas, and John Beatty, eds , *Thinking about Evolution Historical, Philosophical, and Political Perspectives* (New York Cambridge, 2001), pp 309–21

[10] See Gillespie, "Natural Selection for Variances in Offspring Number A New Evolutionary Principle," *American Naturalist*, cxi (1977) 1010–14, and Okasha, p 228

In this way, Okasha seeks to explain modern attitudes toward risk as the product of selection pressures in the past: those of our ancestors that were more circumspect in their actions did better than those that took more risks. This, he submits, gives a better account of the phenomena than the traditional psychological and social scientific theories on offer. However, it turns out that the issues here are significantly more complex than he lets on.

II. ATTITUDES TOWARD RISK: ANOTHER LOOK

In order to see why Okasha's evolutionary argument does not give a completely successful account of human attitudes toward risk, it is necessary to start by reconsidering his conceptualization of the *problem* to be solved. The real issue concerning human attitudes toward risk is not that people are *risk-averse*, but that they are risk-averse *and* risk-loving. More specifically, it is the fact that many people are both risk-loving and risk-averse *at the same time*. Explaining this *nonmonotonicity* in people's attitudes toward risk is the real heart of the problem–making sense only of their risk-*aversion* merely covers half of the ground to be traversed.[11]

A useful illustration of why *this* is the problem to be solved is the following.[12] Many of people across the world take out some form of voluntary *insurance* if it is offered to them (and they are not financially constrained in various ways); however, equally, many people around the world *gamble* in some form or another. Moreover, often the very same people do both: they take out insurance and also play the lottery. Note, however, that taking out insurance only makes sense if the agent is risk-averse: in effect, it replaces a certain loss (the insurance premium) with an uncertain loss that could be either zero or a large amount.[13]

---

[11] This nonmonotonicity has been subject of much speculation in the economic literature ever since Milton Friedman and Leonard Savage's and Daniel Kahneman and Amos Tversky's classic pieces—see Friedman and Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, LVI, 4 (1948): 279–304, and Kahneman and Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, XLVII, 2 (1979): 263–91. More recent discussions of these issues are Roger Hartley and Lisa Farrell, "Can Expected Utility Theory Explain Gambling?," *The American Economic Review*, XCII, 3 (2002): 613–24; Joseph G. Eisenhauer, "How Prevalent Are Friedman-Savage Utility Functions?," *Briefing Notes in Economics*, LXVI (2005): 1–9; Monti, Grant, and Osherson (*op. cit.*); Robert M. Wiseman and Luis R. Gomez-Mejia, "A Behavioral Agency Model of Managerial Risk Taking," *The Academy of Management Review*, XXIII, 1 (1998): 133–53.

[12] See also Friedman and Savage (*op. cit.*), and Hartley and Farrell (*op. cit.*).

[13] The economics of the insurance-industry are complex—the above is just an illustration for expository purposes. For a good introduction to this issue, see Georges

On the other hand, buying a lottery ticket only makes sense on the assumption of being risk-loving it replaces the status quo for a small chance of gain and a large chance of loss (that is, the price of the ticket) [14] Thus, the fact that people frequently act in this way shows that, at least prima facie, what requires explanation is not just why people risk-averse, but why they are *risk-averse and risk-loving*

It is important to note that this is a fundamentally different problem from the one that Okasha tries to solve since explananda generally need not "compose," explaining why people are risk-averse (but not why they are also risk-loving) *and* explaining why people are risk-loving (but not why they are also risk-averse) need not make for an explanation of why people are risk-averse *and* risk-loving [15] That is, the point is not that there is something that is not explained by Okasha's account—it is rather that (at least potentially), *nothing* is This is because, in general, if we want to explain some phenomenon *N*, then it will not be sufficient to start by explaining some other phenomenon *M*—and this is so *even if* it might appear that *M* is somehow "a part of" *N*

As an analogy, imagine two identical metal balls being released in a magnetic field, and that one of them accelerates downwards toward the ground, while the other one moves upwards toward one of the poles of the field In this case, it will not do to explain the behavior of the first ball as due to gravity, and the behavior of the second due to the magnetic field—for what requires explanation is why the first ball behaved in one way and the second in another *in the same circumstances* To be sure, gravity and the magnetic field will need to be cited in this explanation—but their effects cannot be separated out according to what *object* they concern Instead, the *distribution* of the causal factors in *both* cases needs to be explained Why did gravity and electromagnetism combine to yield one effect in one case and another in another?

In the present context, this implies that in order to explain compellingly why people are risk-averse, it may need to be taken into account that they are *also* risk-loving—in short, what requires explanation

---

Dionne and Scott E Harrington, eds , *Foundations of Insurance Economics Readings in Economics and Finance* (New York Kluwer, 1992)

[14] This is so under the assumption that people do not obtain utility from playing the lottery (as opposed to just from receiving its outcomes) Relaxing this assumption makes for another potential psychological explanation of risk-love—see also below

[15] This is also why it matters that the same people are acting in ways that suggest both risk-love and risk-aversion this shows that these are not two disjoint phenomena concerning two different types of psychologies (say), but rather two aspects of the same psychology

is the nature of their attitudes toward risk *in their entirety* This makes for a fundamentally different problem than the one attacked by Okasha

Concerning this more general problem, there are again–in principle—three types of explanations that can be distinguished psychological ones, social scientific ones, and evolutionary ones Again, though, separating the first two is not greatly relevant for present purposes, where distinguishing evolutionary from nonevolutionary explanations is sufficient Equally, it is still the case that the nonevolutionary side—either in the form of a psychological or a social scientific theory—holds the central position in the literature [16]

Among the most prominent explanations of this type are the following (1) It may be that people are risk-loving when there is a chance of winning a significant amount, and risk-averse when it comes to losing small to medium-sized amounts (the Friedman-Savage hypothesis) [17] (2) It may be that they are risk-loving when the alternative to the gamble is a certain *loss*, but risk-averse when it is a certain *gain* (the Kahneman-Tversky hypothesis) [18] (3) There might be external constraints that make it appear *as if* the individuals are risk-averse and risk-loving, when they are in fact risk-neutral [19] Graphically, these hypotheses yield more complex utility functions (compared to Figure 1)—for example, the famous S-shaped ones in Figure 2 (for the Friedman-Savage hypothesis) and Figure 3 (for the Kahneman-Tversky hypothesis)

Now, despite the fact that Okasha has misconstrued the nature of the problem to be solved, he is correct in noting that it is very controversial how successful the nonevolutionary explanations are in dealing with the empirical phenomena For example, there seem to be many cases of risk-loving behavior that concern medium-sized losses (thus contradicting (1) above), concerning certain gains (thus contradicting (2)), and concerning cases where external constraints are absent (thus contradicting (3)) [20] Again, this fact—as well as the intrinsic interest of other explanations—might suggest looking for an evolutionary account as an alternative

However, at this point the two scenarios depart Unlike in Okasha's original conceptualization of the problem, it becomes immediately

---

[16] Again, this strict division is somewhat artificial and more an expository device at this point of the argument See also note 5

[17] See Friedman and Savage (*op cit*), and Hartley and Farrell (*op cit*) The latter piece also makes clear that Rabin and Thaler's objection (*op cit*, note 1, pp 222–23) to this type of utility function appears overly quick

[18] See Kahneman and Tversky (*op cit*)

[19] See, for example, Mahul (*op cit*), and Wiseman and Gomez-Mejia (*op cit*)

[20] For more on this, see the literature cited in note 11 Note also that some of the difficulties here come from the fact that there is empirical support for both (1) and (2), but that the two are at least partly in conflict with one another

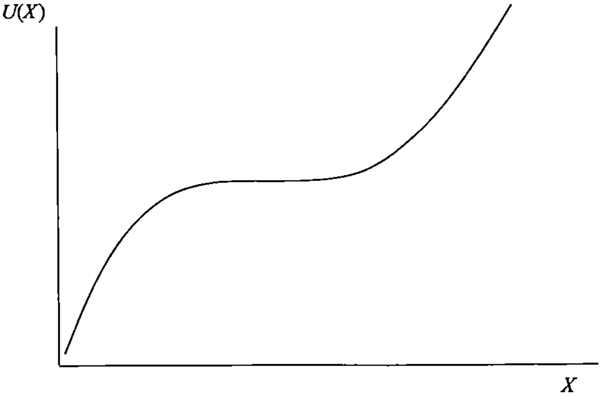


Figure 2

clear that his adaptationist account is in no position whatsoever to say anything about the *general* problem concerning our attitudes toward risk for if there is not only risk-aversion to contend with, but also risk-love, his adaptationist explanation cannot even get off the ground

To see this, note that (as made clear earlier) according to the Gillespie / Okasha measure, the fitness of an organism depends negatively on the variance with which it reproduces This though means that structurally, Okasha's argument as it stands cannot deal with risk-loving behavior, for according to this view, choosing more risky strategies is never adaptive It does not matter if the potential

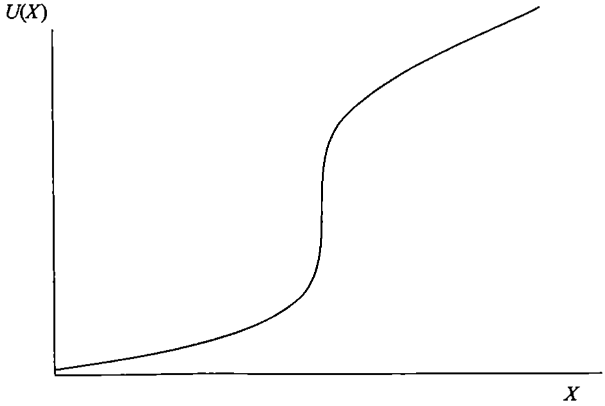


Figure 3

gain is huge and the potential costs are negligible, or if the payoffs involved are small—a risk-averse (that is, low-variance) strategy is always fitter than a risk-loving (that is, high-variance) one.

This might appear slightly counterintuitive at first: it may seem as if adaptationist explanations for risk-love ought to exist as well, since it is relatively easy to construct scenarios according to which strategies with a high variance in reproductive success are fitter than less variable ones.

In order to understand why this is misleading, one has to be careful in distinguishing "variability" from "riskiness": what the Gillespie / Okasha measure of fitness shows is that more variable strategies are never fitter than less variable ones *if the two have the same mean*. If the former has a higher mean than the latter, however, then it *can* be selected after all.[21] Importantly, though, what the definitions of the attitudes toward risk in section I show is that at stake here are instances of *mean-preserving* spreads of reproductive strategies—and it is these that are being selected against according to the Gillespie / Okasha measure of fitness. Furthermore, it is because people *in fact* pursue strategies of exactly this kind that Okasha's adaptationist account cannot provide an explanation of our attitudes toward risk—a fortiori, therefore, it cannot give an explanation that in some way *improves on* the nonevolutionary ones already available.

At this point, there are various paths open along which the ensuing debate could proceed. On the one hand, it may be necessary to accept the conclusion that evolutionary accounts are altogether unhelpful in this context. On the other, it may be possible to construct an extension of Okasha's account that is able to handle the issues after all. Before concluding, it is useful to briefly consider the outlines of such an extension.

This extension begins by noting that a form of Okasha's approach may be salvaged by retreating not from its appeal to the theory of evolution per se, but only from the *adaptationist* version it embodies. That is, one may respond to the above argument by claiming that there *are* useful evolutionary explanations of our attitudes toward risk—it is just that these do not appeal to the attitudes having been adaptive. Instead, these explanations take the attitudes to be the product of *other* forms of evolution: for example, they may be the upshot of entirely nonselective processes (as, for example, drift or mutation), or the by-product of the evolution of some other (psychological) trait.

---

[21] For example, this is what drives the model in Eddie Dekel and Suzanne Scotchmer, "On the Evolution of Attitudes towards Risk in Winner-Take-All Games," *The Journal of Economic Theory*, LXXXVII (1999): 125–43.

Presently, it is enough to note that, while there is no reason to dismiss these sorts of nonadaptationist evolutionary explanations out of hand, a lot of work needs to be done in order for them to be plausible and detailed enough to be insightful. Specific evolutionary models must be proposed that are empirically well grounded, and which have the evolution of risk-love and risk-aversion as their outcome. While not impossible, constructing such models is no trivial task. In short: nonadaptationist evolutionary accounts of our attitudes toward risk have the *potential* to be useful additions to the explanation of this part of our psychology; whether they can actually *fulfil* that potential, though, is yet to be determined.

III. CONCLUSION

I have tried to make clear that Okasha's argument, while immensely valuable for showing the usefulness of an evolutionary perspective on the investigation of human attitudes toward risk, fails to be fully compelling. In particular, I hope to have made clear that his focus on risk-aversion alone cannot do justice to the complexity of people's decisions in the face of uncertainty: since people are risk-loving as well as risk-averse, any explanation of their attitudes toward risk must take both aspects into account. Since Okasha's treatment is unable to do this, it falls short of what is required. While there is a possible reply available to this argument—in the form of a distinction between evolving and adapting—it requires significant further work to be made plausible.

ARMIN W. SCHULZ

University of Wisconsin-Madison

NEW BOOKS

GOWLAND, ANGUS *The Worlds of Renaissance Melancholy Robert Burton in Context* New York Cambridge, 2006 xii, 338 p Cloth $90 00

GRANGER, DAVID A *John Dewey, Robert Pirsig, and the Art of Living Revisioning Aesthetic Education* New York Palgrave Macmillan, 2006 xii, 307 p Cloth $75 00

GRANT, EDWARD *A History of Natural Philosophy From the Ancient World to the Nineteenth Century* New York Cambridge, 2007 xiv, 361 p Cloth $70 00, Paper $24 99

GRAYLING, A C *Truth, Meaning and Realism* New York Continuum, 2007 ix, 173 p Cloth $29 95

GREEN, KEITH *Bertrand Russell, Language and Linguistic Theory* New York Continuum, 2007 ix, 174 p Cloth $120 00

GRIFFIN, DAVID RAY *Whitehead's Radically Different Postmodern Philosophy An Argument for Its Contemporary Relevance* Albany SUNY, 2007 xi, 303 p Cloth $85 00

GRIPPE, EDWARD J *Richard Rorty's New Pragmatism Neither Liberal nor Free* New York Continuum, 2007 ix, 210 p Cloth $110 00

GRISWOLD, CHARLES L *Forgiveness A Philosophical Exploration* New York Cambridge, 2007 xxvi, 242 p Cloth $80 00, paper $21 99

GROSS, DANIEL M *The Secret History of Emotion From Aristotle's Rhetoric to Modern Brain Science* Chicago University Press, 2006 x, 194 p Cloth $35 00

GROSS, MICHAEL L *Bioethics and Armed Conflict Moral Dilemmas of Medicine and War* Cambridge MIT, 2006 xv, 384 p Paper $26 00

GRUMETT, DAVID *Teilhard de Chardin Theology, Humanity and Cosmos* Paris Peeters, 2005 xiv, 314 p Paper $40 00

GSCHWANDTNER, CHRISTINA M *Reading Jean-Luc Marion* Bloomington Indiana UP, 2007 xxiii, 320 p Cloth $65 00, paper $24 95

GUERLAC, SUZANNE *Thinking in Time An Introduction to Henri Bergson* Ithaca, NY Cornell, 2006 xiv, 230 p Cloth $49 95, paper $18 95

GUPTA, ANIL *Empiricism and Experience* New York Oxford, 2006 x, 265 p Cloth $45 00

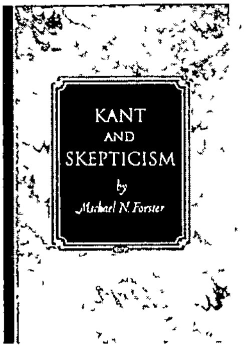
GUTTENPLAN, SAMUEL *Objects of Metaphor* New York Oxford, 2005 305 p Cloth $99 00, paper $35 00

GUYER, PAUL *Kant's Groundwork for the Metaphysics of Morals* New York Continuum, 2007 viii, 186 p Paper $14 95, *Kant's System of Nature and Freedom Selected Essays* New York Oxford, 2005 384 p Cloth $99 00, paper $35 00, *Values of Beauty Historical Essays in Aesthetics* New York Cambridge, 2005 xxi, 359 p Cloth $75 00, paper $27 99

HACKER, P M S *Human Nature The Categorical Framework* Malden, MA Blackwell, 2007 xii, 326 p Cloth $95 00

HAGEN, KURTIS *The Philosophy of Xunzi* Chicago Open Court, 2007 xiv, 176 p Paper $28 95

HAHN, SONGSUK SUSAN *Contradiction in Motion Hegel's Organic Concept of Life and Value* Ithaca, NY Cornell, 2007 xv, 220 p Cloth $49 95

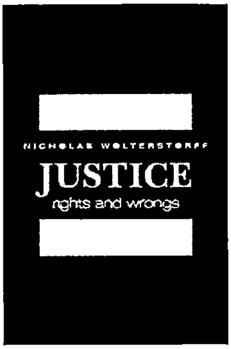


Kant and Skepticism

Michael N. Forster

"This book gives a clear and lucid account of Kant's relation to skepticism, both in its historical development and in its systematic import. Forster connects Kant's views to contemporary discussions even while avoiding the pitfalls of anachronism. The confidence with which he traces the various strands of skepticism in the history of philosophy as well as Kant's responses to them is remarkable."
—Béatrice Longuenesse, New York University

Princeton Monographs in Philosophy
Harry G. Frankfurt, Series Editor
Cloth $29.95 978-0-691-12987-7 Due March



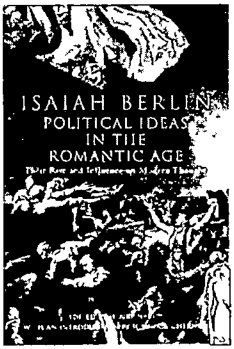
Justice

Rights and Wrongs

Nicholas Wolterstorff

"The work of a first-rate philosopher at the top of his game, this book sets forth a distinctive and challenging theory of justice formulated in explicitly scriptural and Christian terms, yet in conversation with the leading alternatives in the Anglophone world."
—Jean Porter, University of Notre Dame

Cloth $39.50 978-0-691-12967-9



Political Ideas in the Romantic Age

Their Rise and Influence
on Modern Thought

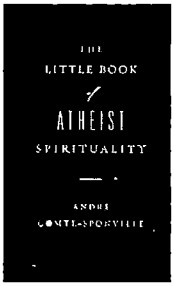
Isaiah Berlin
Edited by Henry Hardy
With an introduction by Joshua L. Cherniss

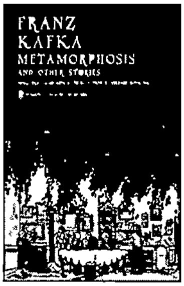
"Indispensable for anyone interested in the history of ideas and the development of liberal thought."
—John Gray, *New York Review of Books*

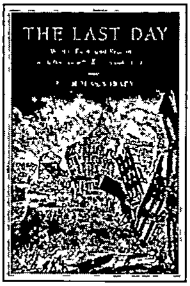
New in Paper $19.95 978-0-691-12695-1 Due April
Not for sale in the Commonwealth (except Canada) and the European Union

PRINCETON UNIVERSITY PRESS
800.777.4726
press.princeton.edu

PHILOSOPHY | PENGUIN GROUP | **PHILOSOPHY** | PENGUIN GROUP







André Comte-Sponville

THE LITTLE BOOK OF ATHEIST SPIRITUALITY

Translated by Nancy Huston
France's preeminent contemporary philosopher presents an exploration of atheism and offers a treatise on a new form of spiritual life "At last, a book that takes the current atheist/believer debate to a much higher, more humane level than any other we know"—Thomas Cathcart and Daniel Klein, authors of *Plato and a Platypus Walk into a Bar—Understanding Philosophy Through Jokes*
Viking • 224 pp • 978-0-670-01847-5 • $19 95

Franz Kafka

THE METAMORPHOSIS and Other Stories

Newly Translated with an Introduction by Michael Hofmann Cover by Sammy Harkham In honor of the 125th anniversary of his birth, this new translation makes Kafka's best-known works available to a new generation of readers
Penguin • 320 pp • 978 0-14-310524-4 • $14 00

Nicholas Shrady

THE LAST DAY

Wrath, Ruin, and Reason in the Great Lisbon Earthquake of 1755

The cataclysmic Lisbon earthquake shook the religious and intellectual foundations of Enlightenment Europe, sparking debate about God's place in human affairs
Viking • 240 pp • 978-0-670-01851-2 • $25 95

Chuang Tzu

THE BOOK OF CHUANG TZU

Translated by Martin Palmer with Elizabeth Breuilly Introduction by Martin Palmer This ancient Chinese classic describes the Tao as a perfect state in which primal, innate nature rules
Penguin Classics
352 pp • 978-0-14-045537-3 • $16 00

John Stuart Mill

ON LIBERTY and THE SUBJECTION OF WOMEN

Edited by Alan Ryan "The present world would be better than it is if Mill's principles were more respected"—Bertrand Russell Introduction and chronology
Penguin Classics
304 pp • 978-0-14-144147-4 • $10 00

Isaiah Berlin, editor

RUSSIAN THINKERS

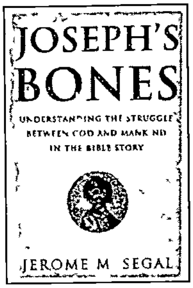
Edited by Henry Hardy
Introduction by Aileen Kelly
This collection of seven essays includes vivid portraits of such remarkable figures as Herzen, Belinsky, Turgenev, and Tolstoy
Penguin • 336 pp • 978-0-14-144220-4 • $17 00

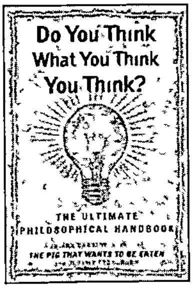
Lucretius

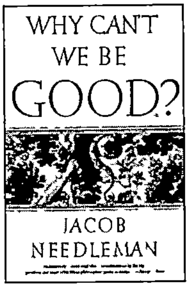
THE NATURE OF THINGS

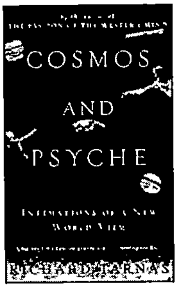
Newly translated by A E Stallings
New Introduction by Richard Jenkyns
This new transaltion of the seminal Epicurean text is exclusive to Penguin Classics
Penguin Classics
336 pp • 978-0-14-04476-5 • $15 00

PENGUIN GROUP (USA) www penguin com/academic


JOSEPH'S BONES
UNDERSTANDING THE STRUGGLE BETWEEN GOD AND MANKIND IN THE BIBLE STORY
JEROME M SEGAL


Do You Think What You Think You Think?
THE ULTIMATE PHILOSOPHICAL HANDBOOK





Jerome Segal

JOSEPH'S BONES

Understanding the Struggle Between God and Mankind in the Bible

A bold reinterpretation of the Hebrew Bible that presents a new understanding of its message through the story of Joseph. "Audacious . . . a work of interpretation that is itself novelistic in its humanistic vision and emotional force." —Naomi Seidman, author of *Faithful Renderings*

Riverhead • 320 pp • 978-1-59448-289-2 • $15.00

Lao Tzu

TAO TE CHING

The New Translation from Tao Te Ching: The Definitive Edition

Translated by Jonathan Star

"It would be hard to find a fresh approach to a text that ranks only behind the Bible as the most widely translated book in the world, but Star achieves that goal. As fascinating to the casual scholar as it is for the serious student."—NAPRA ReView

Tarcher • 112 pp • 978-1-58542-618-8 • $10.00

Julian Baggini & Jeremy Stangroom

DO YOU THINK WHAT YOU THINK YOU THINK?

The founding editors of the *Philosopher's Magazine* present a dozen philosophical quizzes guaranteed to make armchair philosophers uncomfortably shift in their seats.

Plume • 192 pp • 978-0 452-28865-2 • $13.00

Jacob Needleman

WHY CAN'T WE BE GOOD?

"An immensely learned man who is unembarrassed by the big questions." —*Chicago Tribune*

Tarcher • 304 pp • 978-1-58542-620-1 • $14.95

Richard Tarnas

COSMOS AND PSYCHE

Intimations of a New World View

"Breathtaking in the scope and scale of its vision, this extraordinary book shatters our cosmological assumptions."—Christopher Bache, Youngstown State University

Plume • 592 pp • 978-0 452-28859-1 • $18.00

THE ESSENTIAL MARCUS AURELIUS

Cornerstone Edition

Newly Translated with an Introduction by Jacob Needleman and John P. Piazza

A relevant and reliable new translation of the thoughts and aphorisms of the Stoic philosopher and Roman emperor.

Tarcher • 128 pp • 978-1-58542-617-1 • $10.00

Immanuel Kant

CRITIQUE OF PURE REASON

Translated and edited with an new Introduction by Marcus Weigelt. Based on the translation by Max Muller. This seminal text brings together two opposing schools of thought, rationalism and empiricism. New introduction, bibliography and explanatory notes.

Penguin Classics

784 pp • 978-0-14-044747-7 • $20.00

New from OXFORD

The Riddle of Hume's Treatise
Skepticism, Naturalism, and Irreligion
Paul Russell

COGITO?
DESCARTES AND THINKING THE WORLD
JOSEPH ALMOG

Descartes' maxim *Cogito, Ergo Sum* is perhaps the most famous philosophical expression ever coined But how it can be that a thinking human can be both part of the natural world and yet somehow distinct and separate from it? Joseph Almog goes back to the *Meditations* and tries to answer the question
2008 144 pp $35 00

FICTION AND THE WEAVE OF LIFE
JOHN GIBSON

Literary fiction is of crucial importance in human life It is a source of understanding and insight into the nature of the human condition Yet ever since Aristotle, philosophers have struggled to provide a plausible explanation of how this can be the case In *Fiction and the Weave of Life*, John Gibson offers a novel and intriguing account of the relationship between literature and everyday life, and shows how literature can give us an understanding of our world without literally being about our world
2008 234 pp $59 95

THE OXFORD HANDBOOK OF CONTINENTAL PHILOSOPHY
Edited by BRIAN LEITER and MICHAEL ROSEN

The Oxford Handbook of Continental Philosophy is the definitive guide to the major themes of the continental European tradition in philosophy in the nineteenth and twentieth centuries Brian Leiter and Michael Rosen have assembled a stellar group of contributors who provide a thematic treatment of continental philosophy, making this *Handbook* an essential reference point for graduate students and professional academics working on continental philosophy, as well as those with an interest in European literature, the history of ideas, and cultural studies
(Oxford Handbooks in Philosophy)
2008 800 pp $155 00

REVENGE OF THE LIAR
NEW ESSAYS ON THE PARADOX
Edited by JC BEALL

JC Beall presents fourteen new essays and an extensive introduction, which examine the nature of the Liar paradox and its resistance to any attempt to solve it Written by some of the world's leading experts in the field, the papers in this volume will be an important resource for those working in truth studies, philosophical logic, and philosophy of language, as well as those with an interest in formal semantics and metaphysics
2008 296 pp cloth $120 00 paper $49 95

THE RIDDLE OF HUME'S TREATISE
SKEPTICISM, NATURALISM, AND IRRELIGION
PAUL RUSSELL

"Paul Russell's lucid and finely-researched book will enable all of us to read Hume's primary work with fresh understanding, and is a major addition to the scholarly literature"—Terence Penelhum, Professor Emeritus of Religious Studies, University of Calgary
2008 448 pp, 7 b/w line illus $74 00

OXFORD STUDIES IN ANCIENT PHILOSOPHY XXXIII
Edited by DAVID SEDLEY

Oxford Studies in Ancient Philosophy is a volume of original articles on all aspects of ancient philosophy This volume covers a wide chronological range of ancient philosophy, from the Presocratics, Heraclitus and Anaxagoras, to Galen and Aspasius in the second century AD At the core of the volume are five articles on Aristotle
(Oxford Studies in Ancient Philosophy)
2008 376 pp cloth $90 00 paper $50 00

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US To order, please contact customer service at 1 866 445 8685, or visit us online at www oup com/us

Balla
4 11 08

THE JOURNAL OF PHILOSOPHY

VOLUME CV, NUMBER 4
APRIL 2008

page
167 *Indeterminacy and Incomplete Definitions* David Barnett
192 *Supervaluationism and Logical Revisionism* J R G Williams
COMMENTS AND CRITICISM
213 *Understanding the Dimensions of Realization* Thomas W Polger and Lawrence A Shapiro

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Michael J Mooney, *Treasurer,* Liaquat Ahamed, *Other Trustees* Lee Bollinger, Leigh S Cauman, Barbara Gimbel, Kent Greenawalt, Lynn Nesbit, Daniel Shapiro

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2008

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2008 by the Journal of Philosophy, Inc ISSN 0022–362X

THE JOURNAL OF PHILOSOPHY
VOLUME CV, NO. 4, APRIL 2008

INDETERMINACY AND INCOMPLETE DEFINITIONS*

Intuitively, a question is *indeterminate* just in case it is unsettled, not merely epistemically, but metaphysically. We ordinarily ascribe indeterminacy by saying that there is no fact of the matter. Some say, for instance, that there is no fact of the matter how many mountains exist. The topographical facts appear to settle that there are some mountains, but not how many.

While it can seem obvious that certain questions are indeterminate, it is remarkably easy, expanding on arguments by Paul Horwich and Timothy Williamson, to reduce an arbitrary ascription of indeterminacy to absurdity.[1] Suppose, for reductio, that it is indeterminate—that is, metaphysically unsettled—whether Harry is bald. Because it is metaphysically unsettled whether Harry is bald only if it is not metaphysically settled that Harry is bald and not metaphysically settled that Harry is not bald, it is not metaphysically settled that Harry is bald and not metaphysically settled that Harry is not bald. Because it is true that Harry is bald only if it is metaphysically settled that he is bald, and because it is true that Harry is not bald only if it is metaphysically settled that he is not bald, it is not true that Harry is bald and not true that Harry is not bald. Because Harry is bald only if it is true that he is bald, and because Harry is not bald only if it is true that he is not bald, Harry is not bald and Harry is not not bald. This is a contradiction. Hence, our supposition is false. Moreover, because the question of whether Harry is bald was chosen arbitrarily, the argument generalizes: indeterminacy is impossible.[2]

---

* For helpful comments and discussion, I am indebted to David Christensen, Hartry Field, Kit Fine, Christian Lee, the editors of this JOURNAL and, in a special measure, to George Bealer and Adam Pautz.

[1] Horwich, *Truth* (Cambridge: Blackwell, 1990), p. 76; Williamson, *Vagueness* (New York: Routledge, 1994), pp. 187–95.

[2] My argument has a minor advantage over Horwich's and Williamson's. Horwich, *Truth*, p. 76, and Williamson, *Vagueness*, pp. 187–89, argue for bivalence. But, as Williamson

0022-362X/08/0504/167–91

Rejecting this argument requires rejecting at least one of (a) – (c)

(a) it is true that *S* only if it is metaphysically settled that *S*
(b) *S* only if it is true that *S*
(c) modus ponens, conditional contraposability, and reductio ad absurdum are valid rules of inference

Suppose first that we reject (a) We would in particular need to deny that it is true that Harry is bald only if it is metaphysically settled that he is bald But how could it possibly be true that Harry is bald if nothing in the fabric of reality *settles* that he is bald? If the more basic facts about Harry's hair situation, as well as those of his comparison class, together with all the facts about our use of the word 'bald', fail to settle that Harry is bald, then it is hardly plausible that it might be *true* that he is bald Suppose, then, that we instead reject (b) We would in particular need to deny that Harry is bald only if it is true that he is bald But how could Harry possibly be bald if it is not true that he is bald? What more could be required for it to be true that Harry is bald than for him to be bald? Suppose, then, that we instead reject (c) We would in particular need to deny, of at least one of three of the most basic forms of inference, that it is valid when reasoning about even the most mundane of topics The costs of rejecting (c) hardly need emphasis

We are faced with a dilemma take the anti-indeterminacy argument at face value and look for a way to explain away our pro-indeterminacy intuitions, or take our intuitions at face value and give up on one or more of (a) – (c) While many are inclined toward the latter route, a significant few, including Williamson, go the former route

Those of us who go the former route must explain away the pro-indeterminacy intuitions The sources of these intuitions can be di-

---

acknowledges, pp 194–95, the proponent of indeterminacy can accept bivalence by identifying a natural nonepistemic interpretation of 'determinately' according to which, possibly, some utterance is true or false but not determinately true or determinately false Unable to find such an interpretation, Williamson concludes that indeterminacy is impossible But a number of authors at least implicitly suggest such an interpretation x is *determinately* F iff$_{\mathrm{df}}$ it is metaphysically settled that x is F (See James Cargile, "The Sorites Paradox," *British Journal for the Philosophy of Science*, XX (1969) 193–202, Richmond Campbell, "The Sorites Paradox," *Philosophical Studies*, XXVI (1974) 175–91, J A. Burgess, "The Sorites Paradox and Higher-Order Vagueness," *Synthese*, LXXV (1990) 417–74, Horwich, *Truth*, pp 88–90, Vann McGee and Brian McLaughlin, "Distinctions without a Difference," *Southern Journal of Philosophy*, XXXIII (1995, Supplement) 203–51, Harty Field, "Indeterminacy, Degree of Belief, and Excluded Middle," *Noûs*, XXIV (2000) 1–30) This interpretation allows proponents of indeterminacy to escape Horwich's and Williamson's arguments The argument I formulate in the text is designed to block this escape vided into two categories vagueness and nonvagueness related Williamson and others have tried to explain away the vagueness-related intuitions, in part by providing indeterminacy-free accounts of vagueness On the epistemic account, for instance, vagueness is a type of ignorance [3] On the contextualist account, vagueness is a type of shifting context [4] And on the nonreductionist account, vagueness is sui generis [5] In addition to providing a positive indeterminacy-free account of vagueness, proponents of this route need to locate the confusion that gives rise to pro-indeterminacy intuitions [6] Suppose, then, that vagueness-related pro-indeterminacy intuitions can be explained away

The next step toward justifying the anti-indeterminacy route is to explain away nonvagueness-related pro-indeterminacy intuitions I aim in this paper to do this for intuitions related to explicitly incomplete definitions Williamson attempts to explain away these intuitions, but my approach is different [7]

First I state the intuitive case from incomplete definitions to indeterminacy, and I say why Williamson's approach is unsatisfactory (section I) Then I argue that incomplete definitions do not give rise to indeterminacy, and I provide my own explanation of our initial pro-indeterminacy intuitions I do this first for singular terms (section II), then for predicates (section III) I apply our results to two cases first I show how Robert Brandom and Hartry Field go wrong in drawing a pro-indeterminacy conclusion from the incomplete definition of a singular term,[8] then I show how Scott Soames goes wrong in drawing a pro-indeterminacy conclusion from the incomplete definition of a predicate (section IV) [9] Finally, I consider how my view bears on three issues the apparent ubiquity of incomplete definitions in natural language, the Problem of the Many, and the uniqueness problem for arithmetic (section v).

---

[3] See, for example, Roy Sorensen, *Blindspots* (New York Oxford, 1988), and *Vagueness and Contradiction* (New York Oxford, 2001), and Williamson, *Vagueness*

[4] See, for example, Diana Raffman, "Vagueness without Paradox," *Philosophical Review*, CIII (1994) 41–74, and Delia Graff, "Shifting Sands An Interest-Relative Theory of Vagueness," *Philosophical Topics*, XXVIII (2000) 45–81

[5] See my "Is Vagueness Sui Generis?" *Australasian Journal of Philosophy* (forthcoming)

[6] For a sample of attempts, see Williamson, "Imagination, Stipulation, and Vagueness," in *Philosophical Issues, Volume 8 Truth*, E Villanueva, ed (Atascadero, CA Ridgeview, 1997), pp 215–28, Sorenson, *Vagueness and Contradiction*, and my "Is Vagueness Sui Generis?"

[7] Williamson, "Imagination, Stipulation, and Vagueness "

[8] Brandom, "The Significance of Complex Numbers for Frege's Philosophy of Mathematics" (paper read at Proceedings of the Aristotelian Society, 1996), Field, "Indeterminacy, Degree of Belief, and Excluded Middle "

[9] Soames, "Presupposition," in F Guenthner and D Gabbay, eds , *Handbook of Philosophical Logic*, Volume 4, *Topics in Philosophy of Language* (Boston Reidel, 1989), pp 553–616, and *Understanding Truth* (New York Oxford, 1999)

I. THE INTUITIVE CASE FROM INCOMPLETE DEFINITIONS TO INDETERMINACY

Let me introduce the expression 'nice*' by uttering only two sentences:

(1) n is nice* if $n > 15$.
(2) n is not nice* if $n < 15$ (for natural numbers n).[10]

Initially, it seems that 16 is nice*, that 14 is not nice*, and that it is indeterminate whether 15 is nice*.

The intuition that some numbers, including 16, are nice* motivates the standard view of incomplete definitions, on which predicates introduced by incomplete definitions typically succeed in expressing properties (by 'properties' I include relations). On this view, 'is nice*' expresses a property, namely, that of being nice*.

The question then arises: Which property is the property of being nice*? Is it the property of being a natural number greater than 15 or being a natural number greater than 14? The intuition that it is indeterminate whether 15 is nice* motivates the dominant answer to this question, namely, that there is no fact of the matter.[11]

Williamson defends a minority answer.[12] In line with the standard view that 'is nice*' expresses a property, but in opposition to the pro-indeterminacy route, he suggests that being nice* is the property of being a natural number greater than 15. He says, "... given the choice between regarding an atomic predicate as unspecific and so regarding its negation, we generally prefer the latter" (*ibid.*, p. 224). He then proposes the following principle:

> SPEC All other things equal, the application conditions of an atomic predicate are as specific as possible (*ibid.*, p. 224).

---

[10] I derive this example from an example in Kit Fine's "Vagueness, Truth and Logic," *Synthese*, xxx (1975): 265–300, p. 266.

[11] See, for example, Fine, "Vagueness, Truth and Logic"; Horwich, "The Nature of Vagueness," *Philosophy and Phenomenological Research*, LVII (1997): 929–35; Soames, "Presupposition," and *Understanding Truth*; Jamie Tappenden, "The Liar and Sorites Paradoxes: Toward a Unified Treatment," this JOURNAL, XC, 11 (November 1993): 551–77.

For ease of exposition, I formulate the debate in terms of whether certain predicates express properties and, if so, whether there is indeterminacy as to the identities of these properties. The nominalistically inclined reader may reformulate the debate in ontologically more neutral terms, for example, as a debate over whether certain predicates have application conditions and, if so, whether incompleteness in these conditions leads to the possibility of indeterminacy in the application of the predicate.

[12] Williamson, "Imagination, Stipulation, and Vagueness."

Williamson never says exactly what it takes for application conditions to be "as specific as possible " He does however apply SPEC to a case similar to our own From this application, I infer that, by an application of SPEC, 'is nice*' should express the property of being a natural number greater than 15, which in some sense has a more specific range of application than the property of being a natural number greater than 14 According to Williamson, then, our initial pro-indeterminacy intuition arises out of our failure to recognize a key asymmetry between 'nice*' and 'not nice*', namely, SPEC

For three reasons, I cannot accept Williamson's proposal First, the key notion of specificity has not been clearly defined Second, SPEC strikes me as ad hoc Third, SPEC has little or no impact on the force of my initial pro-indeterminacy intuition even after considering SPEC, I still find it strongly intuitive that, if 'is nice*' expresses a property, there is no fact of the matter which property that is

I think the opponent of indeterminacy can do better What is needed is a credible indeterminacy-free account of explicitly incomplete definitions that deflates the force of our initial pro-indeterminacy intuitions In my view, such an account lies in the rejection of the standard view that predicates like 'is nice*' express properties In the next two sections I defend a positive indeterminacy-free rival to the standard view, and I show why our initial pro-indeterminacy intuitions either (i) do not commit us to genuine indeterminacy or (ii) are based on a simple confusion

II INCOMPLETE DEFINITIONS OF SINGULAR TERMS

So far, we have considered an intuitive case for indeterminacy from incompletely defined predicates Now I want to consider a similar case for indeterminacy from incompletely defined singular terms I will show why it fails and explain away the intuition behind it Then, in the next section, I will apply the same strategy to the predicate case

Let me introduce the singular term 'Bitz' by uttering only one sentence

(3) Let Bitz be a resident of New York

Now I ask you to consider whether Bitz is a resident of the U S A and whether Bitz is female Initially, it seems that Bitz *is* a resident of the U S A and that there is no fact of the matter whether Bitz is female This example illustrates the intuitive case for indeterminacy from incompletely defined singular terms

In response to the case, let me begin by drawing your attention to the following uniqueness constraint a singular term refers only if features of its use determine a unique referent for it (that is, only if they determine, of some unique thing, that that thing is its referent.[13]) This constraint has the air of a truism. By definition a singular term purports to refer to a single thing: if it has a referent, it has a unique referent. And it is a platitude about meaning that words have their semantic features determined solely by features of their use (where use is construed broadly, to include both speaker intentions and relations to the environment). Hence, if a singular term refers, features of its use must determine a unique referent for it. (Do not confuse this constraint with the outright rejection of indeterminacy of reference; it does not exclude indeterminacy as to *which* object is so uniquely determined.) We have what appears to be a trivial constraint on reference for singular terms.[14]

Yet in certain contexts we are inclined to act as if there were no such constraint. For instance, even though it is blatantly obvious that the uniqueness constraint on reference for 'Bitz' is unmet, we are inclined to utter 'Bitz is a resident of the U.S.A.' and 'There is no fact of the matter whether Bitz is female'. What explains this inclination?

One hypothesis is that we have stumbled upon a counterexample to the constraint: although nothing in our use of 'Bitz' distinguishes among the millions of candidate referents—the residents of New

---

[13] Why is the parenthetical qualification needed? Because I understand the constraint to be incompatible with supervaluationism. Yet there may be a version of supervaluationism that is compatible with one understanding of the constraint. On this version, on each admissible precisification of our language, for every singular term *t*, the semantic value of 'determines that *t* refers to' is a relation that holds between our use of *t* and exactly one thing; accordingly, the sentence 'a singular term refers only if features of its use determine a unique referent for it' may be deemed *determinately true*. So, to capture my intent, it is not enough to say that a singular term refers only if features of its use determine a unique referent for it. To fully capture the constraint I have in mind, one must say that a singular term refers only if features of its use determine, of some unique thing, that *that* thing is its referent. On this understanding, the supervaluationist must reject the constraint, for it requires an *x* which, on every precisification, satisfies the open sentence 'Features of our use of "Bitz" determine that "Bitz" refers to *x*', and there is no such *x*.

[14] Is the availability of the method of supervaluations evidence against this constraint? No. Of course, *if* the constraint is false, *then* the method of supervaluations might provide a satisfactory semantical account for singular terms (see Bas van Fraassen, "Singular Terms, Truth-Value Gaps, and Free Logic," this JOURNAL, LXIII, 17 (September 15, 1966): 481–95). But the mere existence of the method does not count against the constraint. To the contrary, the apparent triviality of the constraint counts against applying the method. To be sure, there is also initial reason to apply the method, namely, the presence of pro-indeterminacy intuitions. But it is the presence of these intuitions, and not the availability of the formal method, that must be addressed by the opponent of indeterminacy.

For a discussion of the uniqueness requirement on reference for theoretical terms, see David Lewis, "How to Define Theoretical Terms," this JOURNAL, LXVII, 13 (July 9, 1970): 427–46. In opposition to Carnap, Lewis contends that if a theory is either unrealized or multiply realized, its theoretical terms fail to refer.

York—one such candidate is nevertheless the referent of 'Bitz' It is of course indeterminate which such candidate is the referent, and this is why it seems indeterminate whether Bitz is female Still, merely by uttering (3) I succeeded in securing a referent for 'Bitz' If correct, this is a remarkable result not only can a singular term refer if features of its use fail to determine a *unique* referent, it can refer even if features of its use fail to narrow the candidate referents to fewer than 20 million!

Let me suggest a more plausible hypothesis From the perspective of my audience, the most rational way to interpret my utterance of (3) is as an act of *supposition,* rather than as a genuine attempt at reference fixing For if my utterance were a genuine attempt at reference fixing, I would be grossly incompetent I would be under the illusion that I could secure a referent for a singular term simply by narrowing the candidate referents to 20 million My audience might have reason to think that I am incompetent, but not *that* incompetent A reasonable interpretation of my utterance of (3), then, is as an act of *supposing* that 'Bitz' refers to a resident of New York

Given that my utterance was prefaced by, "Let me introduce a new term, 'Bitz'," it is clear that the purpose of my supposition is not to investigate the epistemic possibility that there is a unique person named 'Bitz' who resides in New York Its purpose, rather, is to facilitate discussion of being a resident of New York There is a well-established practice of investigating properties by reasoning *as if* we have concrete instances in mind We prefer concrete instances because we are better at concrete reasoning, we avoid *actual* concrete instances because we do not want to risk tainting our investigation with extraneous detail When a singular term is introduced with no attempt to narrow its candidate referents to one, it is reasonable in many contexts to interpret the introduction as an instance of this practice

Here is a second illustration of the practice

Maxwell What do you know about the living conditions of children in Nigeria?

Emma Let Frib be a five-year-old child in Nigeria Can you guess what the chances are that Frib suffers from malnutrition?

Maxwell I would say at least 30%

Emma Wrong! There is no chance at all For there is no such person as Frib Are you so naïve as to think that I can actually refer to a child in Nigeria just by saying 'Let Frib be a five-year-old child in Nigeria'?

Maxwell I was only cooperating I did not take *you* to be so naïve as to think that you could refer to a child in Nigeria so easily, nor

> did I take you to be out to deceive me I could only figure that you wanted me to go along with you in *supposing* that 'Frib' referred to a five-year-old child in Nigeria I thought that your aim was to teach me something about being a child in Nigeria by reasoning with me *as if* we had a concrete instance in mind Clearly you anticipated that I would reason this way and thus act *as if* I thought there was such a person as Frib, otherwise you would not have set the trap in the first place Of course, you would only have anticipated this had you believed that I was rational and cooperative And so I am flattered

For a third illustration of the practice, suppose that we are discussing mathematics and I say, "Let *a* be an odd number " It would seem misguided for you to interpret my utterance as a genuine attempt at reference fixing It is not as if I am raising my arms to the heavens and asking the gods to make '*a*' refer to some odd number [15] Given that I make no attempt to narrow the candidate referents of '*a*' to one, a rational interpretation of my act is as an act of *supposing* that '*a*' refers to an odd number The supposition is part of the preliminary groundwork for doing mathematics, its purpose is to set the background for reasoning about the property of being odd

Two qualifications are in order First, this proposal is not intended to apply to the mathematical practice of stipulating that a newly introduced function is *undefined* in certain cases Following Frege, we can avoid postulating incompleteness here by taking *undefined* to be a further value (or by means of some other artificial assignment) Second, the proposal is not intended to apply in the context of formal logic In this context, it might be suggested that constants are always treated as successfully referring singular terms On this suggestion, an instance of '[$\phi(a)$]' in an application of existential elimination qualifies, not as a supposition about a vacuous term, '*a*', namely, that it refers to something that satisfies $\phi(x)$, but rather as a supposition about an actual referent of '*a*', namely, that it satisfies $\phi(x)$ I doubt that this is what is going on informally when we say such things as "Let *a* be an odd number " It is not as if we have it in mind that *a* is actually some number, we know not which, and we are going to suppose, of this number, that it is odd To emphasize the point, consider a proof that there is no largest prime which begins, "Let *a* be the largest prime number " Surely we do not think of ourselves as supposing, of some unknown number, that *it* is the largest prime

---

[15] Fine, *Reasoning with Arbitrary Objects* (New York Blackwell, 1985), would interpret my act as an attempt to secure an "arbitrary object" as the referent of '*a*'

Returning to the 'Bitz' example, suppose that I were to make it clear that my act of uttering (3) was *not* an act of supposition: I did not mean to invite my audience to *suppose*, along with me, that 'Bitz' refers to a resident of New York. Rather, I meant to bring it about by my very act of uttering (3) that 'Bitz' *in fact* came to refer to some resident of New York. I suspect that the initial temptations to say that Bitz is a resident of the U.S.A. and that it is indeterminate whether Bitz is female would disappear. I suspect they would be replaced by the temptation to say that 'Bitz' fails to refer and that 'Bitz is a resident of the U.S.A.' and 'Bitz is female' are therefore either false or meaningless (or whatever is the right thing to say on the correct theory of vacuous names). It is too hard to believe that an utterance of (3) might by itself provide a referent for 'Bitz'. One might as well hold that by uttering 'Let "zibot" mean something' an actual meaning might be secured for 'zibot', or that by uttering 'Let "Ball #1" refer to one of the two balls in this urn' an actual referent might be secured for 'Ball #1'. Although it is easy to *suppose* that 'Bitz' refers to a resident of New York, that 'zibot' means something, or that 'Ball #1' refers to one of the two balls in the urn, surely some effort is required to *actually make these states of affairs obtain*.[16]

One might object that I have at most established that 'Bitz' fails to "refer determinately" and that although effort might be required to "secure determinately" a referent for a new term, little or no effort is required to secure a referent for a new term. On this picture, my utterance of 'Let Bitz be a resident of New York' sufficed to secure, but not to "secure determinately," a referent for 'Bitz'. This objection is misplaced. While a consequence of a certain view of indeterminacy is that indeterminacy as to the referent of 'Bitz' is consistent with 'Bitz' referring, just not determinately, my claim is that it is intuitively *not* indeterminate to whom 'Bitz' refers, as it is intuitive—in light of our considerations about supposition—that 'Bitz' fails to refer to anyone.

To establish an indeterminacy of reference one must perform two subtasks, which appear to be in tension. First, one must show that the term in question has properties sufficient to distinguish a *unique* object as its referent. Second, one must show that the question of which object is so distinguished is *indeterminate*. In other words, one must establish that something is more qualified than all others to be the referent of the term *and* that there is no fact of the matter *which* thing is so qualified.[17] For singular terms like ‘Bitz’, ‘Frib’, and ‘Ball #1’, the first subtask cannot be completed.

---

[16] To be sure, we can define a new relation, say, *reference**, as that which obtains between a singular term and a thing iff$_{df}$ the thing is one of possibly many things described or demonstratively picked out as part of the introduction of the term. (Cf. Field's notion of partial reference, "Quine and the Correspondence Theory," *The Philosophical Review*, LXXXIII (1974): 200–28.) And we can say that a singular term *determinately-refers** to a thing iff$_{df}$ it refers* only to that thing. Note: determinately-refers* ≠ determinately refers*. 'Bitz' then refers* to every resident of New York and determinately-refers* to no one. Nothing, however, has been gained in terms of establishing an instance of indeterminacy, for no person is such that it is indeterminate whether 'Bitz' refers* to her.

How, then, should we treat our initial pro-indeterminacy intuition reports? This depends on whether the reports are made from within, or from outside, the scope of the supposition that ‘Bitz’ refers to a resident of New York. Reports made from within can be taken at face value. For they only commit us to genuine indeterminacy *on the supposition* that ‘Bitz’ refers to a resident of New York. And there is good reason to reject this supposition. Pro-indeterminacy reports made outside the scope of the supposition must be rephrased. Given that my audience and I have common knowledge that ‘Bitz’ is a new word that does not refer to anyone, we have common knowledge that the purpose of the supposition is not to explore the epistemic possibility that there is a person named ‘Bitz’ who resides in New York, but rather to investigate the property of being a resident of New York. Given this purpose, my asking a question that is obviously underdetermined by the supposition, for instance whether Bitz is female, is inappropriate. To draw attention to this fact, my audience might disengage from the supposition and misreport the sound intuition *that there is no suppositional fact of the matter* (that is, that the supposition underdetermines the matter) as the unsound intuition *that there is no fact at all of the matter*. Indeed, in light of the internal difficulties with the notion of indeterminacy, it is well advised that we disengage in this way rather than threaten the coherence of our discussion by remaining in the suppositional context and ascribing indeterminacy.

Let me summarize our results so far. The following constraint seems truistic: a singular term refers only if features of its use determine a unique referent for it. Yet, when a singular term is introduced with no effort to narrow its candidate referents to one, we are inclined to act as if this constraint does not exist we act as if the newly introduced term successfully refers, but that there is indeterminacy as to the identity of its referent This behavior can be reconciled with the truistic air of the constraint in light of the fact that the introduction of the term explicitly lacks the degree of determinacy required of an act of reference fixing, we charitably interpret it as an act of supposition, then, in the spirit of cooperation, we suppose that the new term refers Within the scope of this supposition, we are inclined to say things that would normally commit us to indeterminacy and to a counterexample to the uniqueness constraint, however, because these things are said within the scope of the supposition, and because the supposition is not true, no such commitments are made Outside the scope of the supposition, we might misreport the sound intuition *that there is no suppositional fact of the matter* as the unsound intuition *that there is no fact at all of the matter*

---

[17] This task is pursued in a series of articles by Field. On the basis of difficulties resolving issues of reference prior to significant theoretical developments, Field argues that referential indeterminacy has plagued, and likely continues to plague, some of the most central terms of basic science and mathematics. See Field, “Theory Change and the Indeterminacy of Reference,” this JOURNAL, LXX, 14 (August 16, 1973): 462–81, “Quine and the Correspondence Theory,” pp. 200–28, “Are Our Logical and Mathematical Concepts Highly Indeterminate?” *Midwest Studies in Philosophy*, Volume 19 (Minneapolis: Minnesota UP, 1994), pp. 391–429, “Disquotational Truth and Factually Defective Discourse,” *The Philosophical Review*, CIII (1994): 405–52, and “Indeterminacy, Degree of Belief, and Excluded Middle,” *Noûs*, XXIV (2000): 1–30.

III INCOMPLETE DEFINITIONS OF PREDICATES

We began by considering the intuitive case for indeterminacy from incompletely defined predicates, I gave reasons to reject Williamson's treatment of this case Next we considered the intuitive case for indeterminacy from incompletely defined singular terms By focusing our attention on the uniqueness constraint for singular terms, and by explaining away the relevant pro-indeterminacy intuitions, I argued that this case fails

Now I want to return to the case of predicates and argue, in similar fashion, that it fails First I defend a uniqueness constraint on what it takes for a predicate to express a property Then I acknowledge that, in the context of an explicitly incomplete definition, we are inclined to act as if we reject this constraint and accept indeterminacy as to which property is expressed by the new predicate Using the same strategy as above, I argue that these dispositions need not commit us either to a violation of the uniqueness constraint or to an instance of indeterminacy

Here is the constraint a predicate expresses a property only if features of its use determine, of some unique property, that that property is expressed by the predicate[18] To my mind, this constraint has as much initial appeal as the corresponding constraint on reference for singular terms After all, if a predicate '*F*' succeeds in expressing a property, there is a unique property such that '*F*' expresses it, namely, the property of *being F* Because the expressing relation is just another meaning relation, and because words stand in meaning relations only by virtue of how they are used, a predicate expresses a property only if features of its use determine, of some unique property, that that property is expressed by the predicate So it seems to me that the uniqueness constraint for predicates is on a par with that for singular terms

---

[18] To stay neutral on whether there is a meaningful property/concept distinction, I formulate the constraint in terms of properties rather than concepts One who believes in such a distinction may consider a stronger constraint a predicate expresses a property only if features of its use determine, of some unique concept of some unique property, that that concept is expressed by the predicate And one who doubts the existence of properties and/or concepts may reformulate the constraint in ontologically more neutral terms, see note 11

Despite this reasoning, it might be suggested that the uniqueness constraint for predicates is less plausible than that for singular terms and should indeed be rejected The suggestion might be defended as follows Singular terms do not have genuine definitions (at most they have associated reference-fixing descriptions), predicates do Some definitions are partial Partial definitions give rise to incomplete application conditions Because a partially defined predicate *has* application conditions, it expresses a property Because its application conditions are *incomplete,* no property is such that the predicate uniquely expresses it Hence, a predicate might express a property even though no property is such that our use of the predicate uniquely determines, of it, that it is expressed by the predicate Because singular terms do not have genuine definitions, no corresponding argument impugns the uniqueness constraint for them, hence, we should reject the uniqueness constraint for predicates but not for singular terms

This defense fails For if there were such things as partially defined predicates as characterized above, then we could use them, not to *define* singular terms, but to *fix their reference* And if we could do this, we could fix the reference of a singular term without our use determining, of any unique thing, that that thing is the referent of the term

For illustration consider the predicate 'is nice*' Our use of 'nice*' determines that, if 'is nice*' expresses a property, it expresses either the property of being a natural number greater than 15 or the property of being a natural number greater than 14, but nothing in our use determines, of one property or the other, that 'is nice*' expresses it Now suppose that 'is nice*' expresses a property, thus violating the uniqueness constraint for predicates Then we can derive a violation of the uniqueness constraint for singular terms by introducing a new singular term, say, 'Dice*', as follows

(4) Let Dice* be the least nice* number

'Dice*' then refers either to 16 or to 15, yet no features of its use determine, of 16 or of 15, that 'Dice*' refers to it By supposing an arbitrary counterinstance to the uniqueness constraint for predicates, we have derived a counterinstance to the uniqueness constraint for singular terms

There is a systematic method for such derivations It employs a connection between predicates and certain of their nominalizations if a predicate 'F' expresses a property, it expresses the property of being F In other words, 'being F' nominalizes the predicate 'F' it is a singular term that purports to refer to whatever is expressed by 'F' Suppose, then, that 'F' expresses a property even though its use does not determine, of any unique property, that that property is expressed by it Then there is such a property as being F Hence 'being F' refers to a property, yet nothing in its use determines, of a unique property, that that property is its referent We can systematically derive counterinstances to the uniqueness constraint for singular terms given counterinstances to the uniqueness constraint for predicates

The rejection of the uniqueness constraint for predicates leads, then, to the rejection of the uniqueness constraint for singular terms Unless we are willing to admit that a singular term can refer without features of its use narrowing its candidate referents to one, we should not admit that a predicate can express a property without features of its use narrowing the candidate properties to one Our earlier defense of the uniqueness constraint for singular terms thus stands as a defense of the uniqueness constraint for predicates

To be sure, one might draw the conclusion that something must be wrong with our earlier defense of the uniqueness constraint for singular terms But it is incumbent on the proponent of this conclusion to say what exactly is wrong After all, our defense consists of more than just the intuitive appeal of the uniqueness constraint, it includes an explanation of apparent violations of the constraint And there is available a similar explanation of apparent violations of the uniqueness constraint on predicates Why, in the context of an explicitly incomplete definition, are we inclined to act as if we reject the uniqueness constraint for predicates and accept indeterminacy as to which property is expressed by the new predicate? The answer is the same as for singular terms we are conditioned by mathematics and other domains of abstract discourse to treat those acts of word introduction that explicitly lack the degree of determinacy required of successful acts of meaning fixing as acts of supposition

Applied to the original 'nice*' example, we are inclined to interpret (1) and (2) as suppositions In the spirit of cooperation, we *suppose* (1) that n is nice* if $n > 15$ and (2) that n is not nice* if $n < 15$ We are then disposed to accept that 16 is nice*, that 14 is not nice*, and that it is indeterminate whether 15 is nice* Outside the scope of any supposition, this disposition would commit us to a violation of the uniqueness constraint for predicates and to an instance of indeterminacy. However, because it obtains within the scope of a supposition, and because the supposition is not true, no such commitment is made. The supposition is not true because there is no such property as being nice*: the uniqueness constraint for 'is nice*' is not satisfied, hence 'is nice*' fails to express a property.

As in the case of singular terms, rather than ascribing indeterminacy from within the scope of the supposition, we might disengage from the suppositional context and say that there is no fact of the matter whether 15 is nice*, in which case we would be misreporting the sound intuition that there is no *suppositional* fact of the matter—that is, that the supposition fails to settle the matter.

Suppose, then, that I were to make it clear that my utterance of (1) and (2) was *not* an act of supposition, but a genuine attempt at securing a property as the meaning of 'is nice*' and thus as the referent of 'being nice*'. Merely by uttering (1) and (2), I intended to make them true by bringing it about that 'is nice*' *in fact* came to express either the property of being a natural number greater than 15 or the property of being a natural number greater than 14. I suspect that the initial temptations to say that 16 is nice* and that it is indeterminate whether 15 is nice* would disappear. I suspect they would be replaced by the temptation to say that 'is nice' fails to express a property and that '16 is nice*' and '15 is nice*' are therefore either false or meaningless (or whatever is the right thing to say on the correct theory of predicates that fail to express properties).

This concludes my treatment of incomplete definitions.

IV. BRANDOM, FIELD, AND SOAMES ON INDETERMINACY AND INCOMPLETE DEFINITIONS

I have defended what I shall call *the strict view* of incomplete definitions, on which incompletely defined singular terms fail to refer, and incompletely defined predicates fail to express properties. Now I want to apply the strict view to two examples from the literature.

The first concerns singular terms. Robert Brandom and Hartry Field ask us to imagine that a population of speakers separates from ours and that our respective mathematicians independently develop complex-number theory.[19] The only difference is that our mathematicians choose '*i*' and '−*i*', whereas theirs choose '\' and '/', to signify the two square roots of -1 Because the two roots are structurally identical, it is alleged by Brandom and Field to be indeterminate how 'i' and '$-i$' translate '\' and '/' Field says, "It isn't that there is a subtle fact as to 'the correct translation' that we can never know, it is that there is simply no determinate fact of the matter the whole idea of a unique 'correct translation' is misconceived" (*ibid*, p 3) This claim, Field notes, goes hand in hand with the claim that it is indeterminate, of each of the four symbols, to which square root of -1 it refers

---

[19] Brandom, "The Significance of Complex Numbers for Frege's Philosophy of Mathematics"; Field, "Indeterminacy, Degree of Belief, and Excluded Middle."

To me it seems doubtful that 'i' and '$-i$' even purport to refer in the first place (though whether they do is clearly an empirical matter) What a term purports to do depends on the intentions of those who introduce it and, if changes in meaning and reference are possible, of those experts who subsequently use it Given that the experts who use 'i' recognize the impossibility of distinguishing the two square roots of -1, it seems to me unlikely that they intend to use 'i' so as to *refer* to one of the two roots After all, if the experts recognize *their* inability to distinguish between two objects, how could they reasonably expect *their terms* to do it for them? More likely, I think, their use of 'i' involves some sort of linguistic supposition Richard Feynman for instance makes explicit his intention to use 'i' under the supposition that it refers to a specific square root of -1 He says

> Let us suppose that a specific solution of $x^2 = -1$ is called something, we shall call it i, i has the property, by definition, that its square is -1 Someone could write i, but another could say, 'No, I prefer $-i$ My i is minus your i' It *is* just as good a solution, and because the only definition that i has is that $i^2 = -1$, it must be true that any equation we can write is equally true if the sign of i is changed everywhere [20]

Upon considering Brandom's hypothetical scenario, Feynman might either (i) correctly report *from within the scope of his supposition* that there is no fact of the matter whether 'i' and '\' corefer, in which case he would commit to the proposition that the supposition underdetermines whether 'i' and '\' corefer, but not to an instance of indeterminacy, or (ii) misreport *from outside the scope of his supposition* the sound intuition *that there is no suppositional fact of the matter whether 'i' and '\' corefer* as the unsound intuition *that there is no fact of the matter whether 'i' and '\' corefer* On the strict view, speaking outside the supposition, the fact of the matter is that 'i' and '√' do not refer and therefore do not corefer This view is consistent with the fact that practicing mathematicians would never, and should never, deny that $i^2 = -1$, for because the supposition is part of the preliminary groundwork for doing mathematics, practicing mathematicians are always practicing within its scope

---

[20] Feynman, *The Feynman Lectures on Physics*, Volume 1 (New York Addison-Wesley, 1963), p 188

We turn to an example concerning predicates Soames discusses an explicitly incomplete definition similar to that of 'nice*' [21] He says

> Imagine the predicate 'smidget' being introduced into a language by the following semantic stipulation
>
> (i) Any adult human being under three feet in height is a smidget
> (ii) Any adult human being over four feet in height is not a smidget (or is such that it is not the case that he/she is a smidget)
>
> An assertive utterance of 'Jack is a smidget' will convey to one's hearers the information that Jack is an adult under three feet tall and an assertive utterance of 'Jack is not a smidget' will convey the information that Jack is an adult over four feet tall In short, 'smidget' will enter the language as a useful and meaningful predicate
>
> The interesting thing about the predicate is, of course, that the defining conditions for something to be a smidget, and for something to fail to be a smidget, are not jointly exhaustive Adults between three and four feet tall cannot be correctly characterized either as being smidgets or as not being smidgets [22]

There is a temptation to go along with Soames in saying that there is no fact of the matter whether adults between three and four feet tall are smidgets But we need to ask whether this temptation arises in the context of a supposition, and if so whether that supposition is true

Are we merely supposing (i) and (ii), or are we genuinely convinced of their truth? Before accepting (i) and (ii), do we seriously consider whether Soames has the wherewithal to *stipulate* them—to make them true merely by uttering them? Or do we simply go along with him in the spirit of cooperation? Soames prefaces (i) and (ii) by saying, "Imagine the predicate 'smidget' being introduced into a language by the following semantic stipulation " On a reasonable interpretation, he is asking us to *suppose* that 'smidget' is introduced by (i) and (ii) *and* that this act of introduction qualifies as a stipulation It would therefore be reasonable to react to the example by supposing, rather than evaluating, (i) and (ii)

---

[21] Soames, "Presupposition," and *Understanding Truth*

[22] Soames, "Presupposition," pp 584–85

It is true that, within the scope of this supposition, an assertive utterance of 'Jack is a smidget' will convey to one's hearers the information that Jack is an adult under three feet tall, and an assertive utterance of 'Jack is not a smidget' will convey the information that Jack is an adult over four feet tall. It is also true that, within the scope of this supposition, an assertive utterance of 'It is indeterminate whether Jack is a smidget' will convey the information that Jack is an adult between three and four feet tall. And so even if our entire language community were to acknowledge that 'smidget' is introduced by supposition, rather than by stipulation, Soames would be correct in saying that "'smidget' will enter the language as a useful and meaningful predicate." It would enter the language as a *meaningful* predicate, not in the sense that it would actually express a property, but in that acts of predicating it would—in the context of the community-wide supposition—be *meaningful*: they would convey information. All of this is consistent with the impossibility of indeterminacy: if the supposition is false, then ascriptions of indeterminacy within its scope do not carry genuine commitments to indeterminacy.

The question remains: Is the supposition false, or does Soames have the wherewithal to stipulate (i) and (ii)? Soames defends versions of (i) and (ii): "Surely there is no a priori reason why the advantages of introducing a predicate by stipulations of the sort just illustrated must always be outweighed by the potential disadvantages."[23]

If by 'advantages' Soames has in mind *practical* advantages, and if by 'stipulations' he simply has in mind the *acts* of uttering (i) and (ii) regardless of whether they really qualify as stipulations, then we may agree. For we need not deny the potential practical value of introducing a predicate by supposing that it expresses a property without supposing of any particular property that the predicate expresses it.

But if by 'advantages' Soames has in mind features conducive to the success of an act of meaning fixing, and if by 'stipulations' he has in mind acts that rightly qualify as stipulations, then we should disagree. For there *is* an a priori reason why the advantages must always be outweighed by the disadvantages: (i) and (ii) fail to meet the uniqueness constraint on predicates. Contrary to what Soames appears to believe, there are substantive a priori constraints on stipulations involving new words. Speakers cannot just stipulate anything they want about the meanings of new words. I for instance cannot stipulate that 'Bitz' refers to a resident of New York simply by uttering 'Let Bitz be a resident of New York' I can *suppose* that 'Bitz' refers to a resident of New York, but some effort is required to actually make it true Applying our earlier results to Soames's example, we know that if he can stipulate (i) and (ii), then he can also stipulate of a certain singular term that it refers simply by narrowing its candidate referents to a certain uncountable infinity, the relevant term is 'being a smidget', the candidate referents are all properties of the form *being an adult human under k feet in height*—where *k* ranges over the reals in the closed interval [3,4] But if he could do this, then the uniqueness constraints on both predicates and singular terms could be grossly violated And we have not only an a priori argument in defense of these constraints but an explanation as to why, in contexts just like the one Soames creates, we are initially inclined to behave as if we reject the constraints I conclude that the best treatment of Soames's example does not require positing indeterminacy

---

[23] Soames, *Understanding Truth*, p. 164.

V IMPORT FOR A THEORY OF NATURAL AND ARITHMETICAL LANGUAGE

The strict view of incomplete definitions provides for a plausible indeterminacy-free treatment of a certain family of cases, including the 'nice*', 'Bitz', '*i*', and 'smidget' cases Still, one might worry that the view has unacceptable consequences for a theory of natural and arithmetical language In particular, one might worry that it precludes a credible treatment of three issues the apparent ubiquity of incomplete definitions in natural language, the Problem of the Many, and the uniqueness problem for arithmetic

The first worry is that, because natural language is replete with implicit incompleteness of the sort explicit in our 'nice*' example, the strict view leads to a widespread error theory If the correct thing to say about 'is nice*' is that it fails to express a property, then the correct thing to say about most predicates of our language is that they too fail to express properties And if the correct thing to say about 'Bitz' is that it fails to refer, then the correct thing to say about most singular terms of our language is that they too fail to refer But then much of what we say is either false or meaningless, and this is unacceptable

I suspect that this worry depends on the assumption that vagueness is a type of incompleteness in the implicit definitions of our words Vagueness is pervasive in natural language But remember that I began the paper by saying that my goal was to show that incomplete definitions present no obstacle to the anti-indeterminacy route *on the assumption that vagueness presents no obstacle* Moreover, positive incompleteness-free accounts of vagueness are available the indeterminacy-free accounts of vagueness mentioned at the start of our discussion all avoid postulating incompleteness as the source of vagueness Finally, just to whet the reader's appetite, I want to give a phenomenological consideration for thinking that vagueness never in the first place presents as a type of incompleteness

An appearance of incompleteness manifests itself in the form of missing intuitions in certain cases, the word neither seems to apply nor seems not to apply, initially, the number 15 *clearly does not* seem nice* and *clearly does not* seem not nice* By contrast, an appearance of vagueness manifests itself in the form of vagueness in intuition for certain cases, the word *sort of* seems to apply and *sort of* seems not to apply, a person who seems borderline bald *sort of* seems bald and *sort of* seems not bald For emphasis, consider a borderline case of 'seems bald' Suppose that it *sort of* seems to you that Harry is bald Now I ask Does it seem to you that it seems to you that Harry is bald? Your candidate answers are as follows

(i) clearly so
(ii) clearly not
(iii) sort of

(iii) is the obvious answer Unlike incompleteness, vagueness presents as vagueness in intuition Indeed, on the not implausible supposition that meanings of predicates like 'bald' depend on community-wide dispositions to have intuitions about candidate application-conditions, if it is vague whether such a condition suffices for the application of the predicate, it is reasonable to expect vagueness as to whether, upon consideration, the condition would *seem* to us to suffice for the application of the predicate The fact that vagueness presents as vagueness in, rather than as an absence of, intuition is an initial reason to doubt that vagueness is a type of incompleteness I should emphasize, however, that while I am persuaded by this consideration, its success is not required for the thesis of this paper

The second worry is that the strict view precludes a credible solution to Peter Unger's "Problem of the Many "[24] Imagine that from the ground there appears to be a single cloud in the sky I point toward the sky and say, "Let us call that cloud 'McCloud' " On closer examination, what we find is a relatively dense swarm of water droplets The swarm has no sharp boundaries, it gradually fades away as the density of its water droplets decreases The question arises Which collection of water droplets constitutes McCloud? One is inclined to say that there are many equally good candidate collections But if the candidates are equally good candidates for constituting McCloud, then they are equally good candidates for constituting *a* cloud if one of them constitutes a cloud, so do the others Thus, where there initially appeared to be a single cloud, there are either no clouds or many The problem, moreover, has nothing in particular to do with clouds, where there initially appears to be a single table, ship, cat, human, or just about any other ordinary kind of physical object, there are either no such objects or many This is the Problem of the Many

---

[24] Unger, "The Problem of the Many," *Midwest Studies in Philosophy*, v (1980) 411–67

There is a spectrum of positions on the problem, all but two of which are consistent with the strict view of incomplete definitions There is the *nihilistic* position, defended by Unger himself there are no clouds because the concept of a cloud is incoherent (*ibid*) Clearly this position is compatible with the strict view Then there are three positions that appear to require some theory of vagueness, but none in particular First, there is the *restricted composition* position, defended by Ned Markosian one and only one of the many sets of water droplets composes anything at all, and that thing is a cloud[25] Second, there is the *relative identity* position, defended by P T Geach the many sets of water droplets are distinct collections of water droplets but one and the same cloud[26] Third, there is the *constitution is not identity* position, defended by E J Lowe and Mark Johnston the many sets of water droplets compose many distinct collections of water droplets, one and only one of which constitutes a cloud[27] None of these three positions qualifies as a complete solution, for each leaves open a crucial question The restricted-composition position leaves open *which* set of water droplets composes a cloud, the relative-identity position leaves open *which* collections of water droplets are one and the same cloud, and the constitution-is-not-identity position leaves open *which* collection of water droplets constitutes a cloud To each of the three questions, a plausible answer is *that it is vague* Insofar as incompleteness-free analyses of vagueness are available, these three positions are compatible with the strict view

Two positions are incompatible with the strict view The first invokes a particular theory of vagueness By it

(5) Vagueness is a type of incompleteness in the implicit definitions of words

(6) The proper semantics of incompletely defined words is supervaluational

---

[25] Markosian, "Brutal Composition," *Philosophical Studies*, xcii (1998) 211–49

[26] Geach, *Reference and Generality* (Ithaca Cornell, 1980)

[27] Lowe, "The Paradox of the 1,001 Cats," *Analysis*, xlii (1982) 27–30, Johnston, "Constitution Is Not Identity," *Mind*, ci (1992) 89–105

By the method of supervaluations

(7) Clearly, there is one and only one cloud
(8) Clearly, McCloud is this cloud
(9) It is vague which collection of droplets constitutes McCloud [28]

Evaluated simply as a proposal for solving the Problem of the Many, and not as a theory of vagueness, this proposal is appealing because it allows us to maintain (7) – (9) But so do all of the aforementioned indeterminacy-free alternatives to treating vagueness as a type of incompleteness And so, debates over the nature of vagueness aside, it is not a problem for the strict view that it precludes this particular proposal for solving the Problem of the Many

David Lewis proposes a second solution that is incompatible with the strict view [29] He maintains (10) – (13)

(10) Each of the many relevant collections of water droplets constitutes a cloud

Thus, strictly speaking, where there initially appeared to be a single cloud, there are actually many However, because the collections of droplets largely overlap one another, they are "almost identical," and for the purposes of ordinary conversation, this is sufficient for us to count them as one cloud

(11) There is one and only one cloud

So far, the proposal is compatible with the strict view The question arises, however, whether 'McCloud' refers to a cloud, and if so, which one The proposal continues

(12) 'McCloud' refers to a cloud
(13) It is indeterminate to which cloud 'McCloud' refers [30]

Now the proposal is incompatible with the strict view (insofar as it requires a violation of the uniqueness constraint on reference for singular terms) But I do not see this is as a problem for the strict view I pointed at what appeared to be a single cloud and said, "Let us call that cloud 'McCloud'." Supposing now that what appeared to be a single cloud was, strictly speaking, a massive group of partly overlapping clouds, it seems to my mind absurd to think that I successfully referred to one of those clouds. Surely the initial appeal of (12) is contingent on the appearance of a *unique* cloud in the sky. Once we accept (10) and thus accept that, strictly speaking, there was no unique cloud at which I pointed, the initial appeal of (12) disappears. So, rather than presenting a problem for the strict view of incomplete definitions, Lewis's proposal faces the problem of conflicting with an intuitive core principle of the strict view, namely, the uniqueness constraint on reference for singular terms.

---

[28] A supervaluationist solution is defended by Brian Weatherson, "The Problem of the Many," *The Stanford Encyclopedia of Philosophy*, Edward Zalta, ed (Winter 2005 Edition, http //plato stanford edu/archives/win2005/entries/problem-of-many/) and, to some extent, by Lewis, "Many, but Almost One," in John Bacon, Keith Campbell, and Lloyd Reinhardt, eds , *Ontology, Causality and Mind Essays in Honour of D M Armstrong* (New York Cambridge, 1993), pp 164–82, and by McGee and McLaughlin, "The Lessons of the Many," *Philosophical Topics*, xxviii (2000) 129–51

[29] Lewis, "Many, but Almost One "

[30] Lewis, "Many, but Almost One," p 36, suggests a supervaluational treatment of denoting expressions such as 'McCloud'

I doubt, then, that the Problem of the Many poses a serious difficulty for the strict view.

The third worry is that the strict view precludes a credible treatment of the uniqueness problem for arithmetic. The problem is that our arithmetical practice appears to underdetermine the identities of the natural numbers. Because the Peano axioms are satisfied by every countable progression of elements of any kind, they alone do not appear rich enough to determine the identities of the numbers, and it is not obvious what other features of our practice might help. The problem is a special underdetermination problem for two reasons first, standard causal accounts of reference plausibly do not apply to the case of numbers, second, it is hard to see what epistemic access we could have to numbers. There are various versions of the problem. Paul Benacerraf puts it in terms of questions of cross-theory identification,[31] for instance, is 1 identical to {{0}} or to {0,{0}}? Harold Hodes puts it in terms of the question of what, if anything, in our use of arithmetical terms could distinguish unique referents for them.[32]

I will not attempt an exhaustive survey of positions on this problem. For present purposes, it will suffice to consider a few broad categories and show that a number of credible options are compatible with the strict view.[33]

First consider positions on which the numerals do not refer. On this view, speaking strictly, literally, and outside the scope of all suppositions, there are no numbers. Proponents of this view may interpret the Peano axioms as suppositions, or as stipulations in a story of fiction, as Field and Steven Wagner do[34] Speaking outside the supposition/fiction, there are no such things as numbers Speaking within the scope of the supposition/fiction, there are numbers, but their identities are indeterminate Because this ascription of indeterminacy is made within the scope of the relevant supposition/fiction, it does not carry a commitment to genuine indeterminacy, but only to the underdetermination of the question by the supposition/fiction Clearly this view is compatible with the strict view

---

[31] Benacerraf, "What Numbers Could Not Be," *Philosophical Review*, LXXIV (1965) 47–73.

[32] Hodes, "Logicism and the Ontological Commitments of Arithmetic," this JOURNAL, LXXXI, 3 (March 1984) 123–49.

[33] For ease of presentation, I will blur certain distinctions among views in a given category.

Next consider positions on which the numerals do refer These positions divide over whether our use of the numerals determines a unique set of referents for them

There are at least four sorts of views on which our use succeeds in determining a unique set of referents First, there are the nonreductionist Platonist views—held perhaps by Kurt Gödel[35] and perhaps even by certain formalists, such as David Hilbert[36]—which appeal to a special mode of access to a unique set of primitive entities (namely, the natural numbers) that serve as the referents of the numerals[37] Second, there is the nonreductionist view of Kit Fine, which appeals to the method of postulation to secure a unique set of primitive entities as the referents of the numerals[38] Third, there are the reductionist views of Gottlob Frege, Bertrand Russell, Russell and Alfred Whitehead, Mark Steiner, George Bealer, and others, which appeal to features of use beyond the Peano axioms—for instance, ordinary-language idioms of cardinality and perhaps other categorial considerations—to provide evidence for the reduction of the referents of the numerals to unique reducing entities (for example, sets, properties) [39] Fourth, there are the structuralist views of Benacerraf, Michael Resnik, Stewart Shapiro, and others, which appeal simply to Peano's axioms to identify a unique structure whose "places" or "positions" serve as the referents of the numerals [40] All four of these views are compatible with the strict view, for none posits a relevant incompleteness in the definitions of the numerals

---

[34] Field, *Science without Numbers* (New York Blackwell, 1980), Wagner, "Arithmetical Fiction," *Pacific Philosophical Quarterly*, LXIII (1982) 255–69

[35] Gödel, "Russell's Mathematical Logic," in Paul A Schilpp, ed , *The Philosophy of Bertrand Russell* (Evanston, IL Northwestern, 1944), pp 125–53

[36] Hilbert, "Über das Unendliche," English translation in J van Heijenoort, ed , *From Frege to Godel A Source Book in Mathematical Logic, 1897–1931* (Cambridge Harvard, 1967), pp 367–92, originally published in 1926

[37] By 'Platonist view' I mean to pick out a position readily attributed to Plato by those who are willing to set aside certain scholarly issues Whether Plato actually held that numbers were primitive entities is unclear, for discussion, see Anders Wedberg, *Plato's Philosophy of Mathematics* (Stockholm, Sweden Almqvist and Wiksell, 1955), Julia Annas, *Aristotle's Metaphysics Books M and N* (New York Oxford, 1976), Introduction, C D C Reeve, *Philosopher-Kings The Argument of Plato's Republic* (Princeton University Press, 1988)

[38] Fine, "The Limits of Abstraction," in M Schirn, ed , *The Philosophy of Mathematics Today Proceedings of an International Congress in Munich* (New York Oxford, 1988), pp 503–629, *The Limits of Abstraction* (New York Oxford, 2002), and "Our Knowledge of Mathematical Objects," in John Hawthorne and Tamar Gendler, eds , *Oxford Studies in Epistemology Volume 1* (New York Oxford, 2006), pp 89–110

Next consider views on which the numerals successfully refer, but their use fails to determine a unique set of referents for them Here one might hold, for instance, that it is indeterminate which of the rival set-theoretic reductions contemplated by Benacerraf is correct [41] Or one might endorse a more extreme indeterminacy, as Field suggests, on which it is indeterminate, not only which sets numbers are, should they be sets, but what general sorts of things numbers are in the first place primitive entities, concepts, properties, sets, or positions in ω-sequences [42] Both positions are incompatible with the strict view, for both entail that the numerals successfully refer even though nothing in our use of them determines a unique set of referents for them Again, I do not see this as a problem for the strict view To the contrary, it appears to be a problem for the two positions If the choice among the candidate sets of referents is indeed arbitrary, then it seems wrong to conclude that one of the sets is in fact the correct candidate, regardless of whether one is then forced to add the further claim that it is indeterminate which such set is the correct candidate On the supposition that our use underdetermines the referents of numerals, the sensible thing to say is that the numerals fail to refer To be sure, they might still bear reference-like relations to the candidate referents [43] But it seems wrong to maintain that they genuinely refer Field himself seems to acknowledge as much in his move away from his original suggestion (*ibid*), and toward fictionalism [44] And let us not forget that each of the four aforementioned indeterminacy-free alternatives is still very much alive as far as anyone has shown

---

[39] Frege, *Die Grundlagen der Arithmetik eine logisch-mathematische Untersuchung uber den Begriff der Zahl* (Breslau W Koebner, 1884), translated as *The Foundations of Arithmetic A Logico-Mathematical Enquiry into the Concept of Number* by J L Austin (Oxford, UK Blackwell, 1974, second revised edition), Russell, *The Principles of Mathematics* (New York Cambridge, 1903), Russell and Whitehead, *Principia Mathematica*, 3 volumes (New York Cambridge, 1910, 1912, 1913), Steiner, *Mathematical Knowledge* (Ithaca Cornell, 1975), Bealer, *Quality and Concept* (New York Oxford, 1982)

[40] Benacerraf, "What Numbers Could Not Be", Resnik, *Mathematics as a Science of Patterns* (New York Oxford, 1997), Shapiro, *Philosophy of Mathematics Structure and Ontology* (New York Oxford, 1997)

[41] Benacerraf, "What Numbers Could Not Be "

[42] Field, "Quine and the Correspondence Theory "

[43] Cf Field's notion of partial reference in "Quine and the Correspondence Theory "

[44] Field, *Science without Numbers*, and *Realism, Mathematics, and Modality* (New York Blackwell, 1989), Introduction

I conclude that the uniqueness problem for arithmetic poses no serious difficulty for the strict view of incomplete definitions

VI CONCLUSION

I have argued that the standard view of incomplete definitions is incorrect singular terms introduced by incomplete reference-fixing acts do not refer, and predicates introduced by incomplete definitions do not express properties In place of the standard view, I have developed an indeterminacy-free rival the strict view of incomplete definitions My defense of the strict view includes a strategy for explaining away our initial pro-indeterminacy intuitions If my account is sound, incomplete definitions present no obstacle to taking our opening anti-indeterminacy argument at face value

One could not at this stage show decisively that indeterminacy can be avoided universally But if the hardest challenges can be credibly met, we are forced to take a serious look at the view that indeterminacy is impossible

DAVID BARNETT

University of Colorado at Boulder

SUPERVALUATIONISM AND LOGICAL REVISIONISM*

In the literature on supervaluationism, a central source of concern has been the acceptability, or otherwise, of its alleged logical revisionism Timothy Williamson claims that supervaluationism gives rise to

> breakdowns of the classical rules of contraposition, conditional proof, argument by cases and *reduction ad absurdum* in the supervaluationist logic of 'definitely' [1]

Williamson is unhappy with such revisionism

> Conditional proof, argument by cases and reduction ad absurdum play a vital role in systems of natural deduction, the formal systems closest to our informal deductions Supervaluationists have naturally tried to use their semantic apparatus to explain other locutions If their attempts succeed, our language will be riddled with counterexamples to the four rules (*ibid*)

Others, accepting the case for logical revisionism, have argued that the upshot is unobjectionable Thus Rosanna Keefe

> A number of commentators have emphasized how supervaluationist logic fails to preserve certain rules of inference or classical principles about logical consequence
>
> How important is the failure of certain classical principles governing logical consequence? My reply is that the described features of supervaluationism are acceptable [2]

The case for revisionism that these authors present depends on the presence, in the language at hand, of the supervaluationist notion 'Definitely' (the '*D*' operator) [3] Roughly, 'Definitely *p*' says that, no matter how we sharpen the indeterminacy in our language, *p* always holds In this, it is an object-language reflection of the supervaluationist's notion of truth—"supertruth" the idea being that '*p*' is supertrue if it is true no matter how we sharpen our language In a supervalua-

---

*Thanks to the members of *Arche* centre in St Andrews, and the Centre for Metaphysics and Mind in Leeds, for discussion Thanks in particular to JC Beall, Ross Cameron, Andrew McGonigal, Joseph Melia, Greg Restall, Achille Varzi, and Elia Zardini for discussion Many thanks also to Jamie Stark for typesetting work

[1] Williamson, *Vagueness* (New York Routledge, 1994), pp 151–52

[2] Keefe, *Theories of Vagueness* (New York Cambridge, 2000), pp 176–78

[3] Though as both authors emphasize, related notions would also lead to complaints

0022-362X/08/0504/192–212

tional language without the *D*-operator and its relatives, there is no special threat to the classical modes of inference Once it is added, it is alleged that the following results hold, providing counterexamples to the respective classical mode of inference[4]

Contraposition

- $p \models_{sv} Dp$
- $\neg Dp \not\models_{sv} \neg p$

Conditional proof

- $p \models_{sv} Dp$
- $\not\models_{sv} p \supset Dp$

Argument by cases

- $p \models_{sv} Dp \vee D\neg p$
- $\neg p \models_{sv} Dp \vee D\neg p$
- $\not\models_{sv} Dp \vee D\neg p$

Reductio

- $p \wedge \neg Dp \models_{sv} \bot$
- $\not\models \neg(p \wedge \neg Dp)$

Here, for example, is Keefe on contraposition

> In any specification-space where *A* is super-true, *DA* is also super-true since *DA* is defined as true whenever *A* is true on all specifications But it is not typically the case that $\neg DA \models_{sv} \neg A$ in a specification space where *A* is true on some specifications and false in others, $\neg DA$ is super-true, while $\neg A$ is not[5]

Below, I argue that these results do not hold if the supervaluationist's framework is properly chosen My positive case against revisionism will rest on a deniable premise concerning logicality I strengthen the result, therefore, by showing that even if one denies my premise, and (further) can sustain the counterexamples *stricto sensu*, then the "revisionism" that ensues is not damaging, indeed, whether the failure of these modes of inference really deserves the name 'revisionism' becomes a terminological issue

I RESPONSES

I distinguish two ways for the supervaluationist to respond to the alleged counterexamples to classical modes of inference given above

---

[4] The following are taken from Williamson, *op cit*

[5] Keefe, *op cit*, p 176

The first is to claim *contra* Williamson and Keefe, that supervaluational consequence is thoroughgoingly classical, at least in the standard, single-conclusion setting The second is to accept that the examples arise and, with Keefe, to argue that the revisionism induced is not objectionable It is the first line of response that concerns me at present

I can think of three ways of making the case that supervaluationism requires no departure from classical logic

(1) In order to deny that the examples arise, one might give a nonstandard treatment of the connectives involved, $\supset$,$\vee$ and so forth So one would make a case that, properly understood, $\models_{sv} p \supset Dp$ holds, so that the validity of $p \models_{sv} Dp$ does not lead to a failure of conditional proof

(2) One might try to undermine the case by characterizing $\models_{sv}$ in such a way that results such as $p \models_{sv} Dp$ does not hold, so that we can retain the standard treatment of connectives, and still not fall into revisionism

(3) One might make a case that the examples given above are not revisions of classical logic at all, because classical logic fails to sustain the relevant inferences This would involve claiming that conditional proof, argument by cases and the rest are not universally valid, even for the thoroughgoing classicist

I will focus here on (2) I argue that the natural generalization of the classical characterization of logical consequence will give a version of $\models_{sv}$ that does not lead to departures from classical logic indeed, I will argue that none of the cases that Williamson and Keefe cite provide counterexamples to the rules they mention Furthermore, I do this while accepting much of Williamson's setting[6] in particular, his rejection of "local" characterizations of consequence in favor of "global" characterizations

II THE SETTING

Let me begin by outlining the treatment of consequence I favor With Williamson I shall characterize consequence model-theoretically The first challenge therefore is to say what a supervaluationist model looks like

A supervaluationist model structure for a language *L* will consist, at minimum, of a domain of individuals *D*, a set of "delineations" Δ, and appropriate accessibility relations on the domain In addition, there will be an interpretation function *f*, which will assign to expressions classical extensions. Formally, the model-structure so characterized is exactly analogous to that appropriate to a possible-worlds treatment of a modal language.

---

[6] Williamson gives arguments for the revisionary consequences I do not dispute that the results follow from the framework for supervaluationism that he sets up—I dispute elements of that framework on which the cogency of his arguments turn

The crucial difference between my setting, and those of Williamson and Keefe, is that I assume that within the model structure of the intended model there will be delineations that are "extreme"—relative to which a 6′8″ man is short, for example. There are several reasons for wanting to have such delineations within our model structure. Consider the following supervaluationist treatment of comparatives.[7]

> '*A* is *F*-er than *B*' is true at delineation *d* iff the set of delineations where '*A* is *F*' is true is a proper superset of the set of delineations where '*B* is *F*' is true.

This will be untenable unless we have available extreme delineations. For otherwise, the set of delineations making-true '*A* is tall' (where *A* is 6′8″) will be the same as that making-true '*B* is tall' (where *B* is 6′10″). On the treatment of comparatives given above, this will mean that '*B* is taller than *A*' will be declared false, which is absurd.[8]

Nevertheless, we want 'a 6′8″ man is tall' to be true, and 'a 6′8″ man is short' to be false, on the intended model. Due to the presence of extreme delineations, we cannot characterize supertruth in the simplest way, that is, saying that *S* is supertrue iff on every delineation *d*, *S* is true relative to *d*. Some extra machinery is called for.

We first require models to pick out a subset of the delineations—a subset we will call the *sharpenings*. *S* will be supertrue (at a model) if it is true relative to each of the sharpenings of that model. A sentence will be supertrue *simpliciter* if it is supertrue at the *intended* model.

Second, we introduce an accessibility relation, '*S*-access', on the space of delineations of the model structure, and let '*Dp*' be true at a delineation *d* in a model if '*p*' is true at all delineations *S*-accessed by

---

[7] Compare D. K. Lewis, "General Semantics," *Synthese*, xxii (1970): 18–67, reprinted with postscript in Lewis, *Philosophical Papers I* (New York: Oxford, 1983), pp. 189–229; J. A. W. Kamp, "Two Theories about Adjectives," in Edward Keenan, ed., *Formal Semantics of Natural Language* (New York: Cambridge, 1975), pp. 123–55, reprinted in Steven Davis and Brendan Gillion, eds., *Semantics: A Reader* (New York: Oxford, 2004), pp. 541–62.

[8] Williamson and Keefe cite such cases as objections to the treatment of comparatives, under their assumption that all delineations are nonextreme.

A second example: my desk is definitely flat. But in some extreme sense, it is not flat—it is less flat than an oil slick, for example. To give a treatment on which 'this is definitely flat, but in some extreme sense, it is not flat' will come out true, we need to appeal to extreme delineations "accessed" by 'in some extreme sense'.

Yet another reason for wanting such delineations is the need to treat higher-order vagueness within the supervaluationist setting. See Williamson, *op. cit.*, sec. 5.7.

CALCUTTA

d On the intended model, the set of sharpenings will *S*-access each other[9] Our models will therefore take the form

$$m = <D_m, \Delta_m, S_m, f_m, s_m>$$

where D_m is the domain, Δ_m the set of all (extreme and nonextreme) delineations, S_m an accessibility relation, and s_m a subset of the delineations—the sharpenings of the model[10]

Once we have dropped the idea that delineations correspond to the intuitive notion of "precisification of the language," there seems no reason to put *any* constraint what delineations there can be We may as well let there by a delineation corresponding to *any* assignment of semantic values to nonlogical terms Only some among these arbitrary delineations will be sharpenings, and *S*-access each other on the intended interpretation[11] I assume in what follows, therefore, that the delineations in the intended model of our language are arbitrary, though only some among them are sharpenings[12]

Given this setting, we can then distinguish two forms of consequence The first is *local* consequence

$\Gamma \models_{\text{local}} \Phi$ iff
On all models *m*, and all $d \in \Delta_m$, f_m makes Γ true relative to *d* only if f_m makes Φ true relative to *d*

Williamson rejects this characterization on behalf of supervaluationists He takes it that consequence should be characterized in terms of *truth* preservation under arbitrary reinterpretations For standard supervaluationists, then, it should be characterized in terms of supertruth preservation But given this, local validity looks suspect

---

[9] Another constraint on *S* will be, presumable, that it is reflexive More constraints will presumably flow from an adequate account of higher-order vagueness

[10] We can, of course, *introduce* an operator 'D^*' such that D^*p is true iff *p* is true on all delineations *whatsoever* The interest of such a notion is questionable, particularly in a framework involving "arbitrary delineations" rather than merely extreme ones It also will not threaten any of the logical results later in the essay the proof given in the appendix shows that the logic of this operator will be S5, just as is with the analogous introduction of modal operators Thanks to an anonymous referee for raising this issue

[11] It may be that there is semantic work for the notion of an "extreme" but not arbitrary delineation For example, running the story about comparatives above with arbitrary delineations in place of extreme delineations would not be plausible there are *arbitrary* delineations where 'tall' holds of all and only 6′10″ men Thus, the set of arbitrary delineations where a 6′8″ man falls under 'tall' will not be a *proper* subset of the arbitrary delineations where a 6′10″ man falls under 'tall' However, for such semantic purposes, we need only introduce a new accessibility relation, *E*-access, whereby only the delineations I earlier called "extreme" *E*-access each other We would then reformulate the account of comparatives given earlier '*x* is taller than *y*' is true at an (arbitrary) delineation *d* iff the set of delineations *E*-accessible from *d* where '*x* is tall' is true is a subset of those *E*-accessible from *d* where '*y* is tall' is true

[12] Though this seems the natural setting to me, I do not believe the arguments below depend on accepting it as opposed to one where a delineation is at worst "extreme " All that needs to be maintained is that *truth* is to be analyzed not as *truth on all delineations* but as *truth on all sharpenings*

> The problem for supervaluationists is that supertruth plays no role in the definition of local validity Yet they identify truth with supertruth, since validity is necessary preservation of truth, they should identify it with necessary preservation of supertruth That amounts to an alternative definition [13]

This alternative is a *global* consequence[14]

$\Gamma \models_{\text{global}} \Phi$ iff
On all models *m*, Γ is supertrue-at-*m* only if Φ is supertrue-at-*m*

The notion of supertruth—truth at all sharpenings—is the central notion used to define supertruth on a model Hence global consequence meets Williamson's constraints

One might wonder about whether the intuitive connection between validity and necessary truth preservation must be maintained at all costs Would anything go wrong if the supervaluationist appealed to local consequence? Things would indeed go wrong in the following section I point to considerations showing that $\models_{\text{global}}$ is indeed the proper explication of $\models_{sv}$

III REASONS TO REJECT LOCAL CONSEQUENCE

There has been some debate in the literature about the relative merits of defining consequence in local and global ways I side with those would take *global* characterization to be the proper one The defenders of local consequence claim it avoids the revisionism that allegedly follows from the global characterization But I will argue that the local consequence follows the classical paradigm *too closely*[15]

---

[13] Williamson, *op cit*, p 148

[14] Williamson does not mention the relativization of truth to a model in the above But I take it that he will wish to generalize over models in the final characterization, on pain of admitting water = H_2O as a logical validity

If we added into our models a set of possible worlds, and a specification of one among these as "actual," then generalizing over all models would amount to requiring *necessary* truth preservation, as Williamson requires

[15] Delia Fara argues that supervaluationists have independent reason to give up rules such as contraposition, and so on (Fara, "Gap Principles, Penumbral Consequence, and Infinitely Higher-order Vagueness," published under the name "Delia Graff" in JC Beall, ed , *Liars and Heaps New Essays on Paradox* (New York Oxford, 2003), pp 195–221) I think her arguments are resistable, but I will not defend this here The present section, though, makes an analogous move the supervaluationist should not buy into a consequence relation that is extensionally equivalent to the full (multi-conclusion) classical relation

The case is easiest to make by considering one of the most persuasive ways of introducing a sorites paradox Take a paradigmatically red color-patch A, and a paradigmatically blue color patch Z Ask one's audience to endorse both the following

(1) A is red
(2) Z is not red

Now present the audience with a series of conjunctive statements concerning a sorites series $A = A_1, A_2, \;\; ,A_n = Z$, and ask them to *reject* each of them

- A_2 is red $\wedge$ A_2 is not red
- A_2 is red $\wedge$ A_3 is not red
-
- $A_{n\text{-}1}$ is red $\wedge$ Z is not red

Rejecting each of these premises is *prima facie* highly plausible It is one of the advantages of supervaluationism that it can preserve this intuition none of the conjunctive statements is supertrue (at best they are neither true nor false) And given that one should reject p if p fails to be true, each of the conjunctions is rejectable [16]

Often, the next move is to invite one's audience to move from the *rejection* of each of the conjunctions to that *acceptance* of their negations Granted (1) and (2), and each of the negated conjunctions, standard (single-conclusion) classical logic allows one to derive a contradiction The supervaluationist resists the sorites paradox by resisting the move from *rejecting* p to *accepting* $\neg p$ for in case p is neither true nor false, then we should reject both p and $\neg p$, and accept neither

However, there is a standard way of extending logics so as to capture the logical relationship in this case more directly One way of expressing the practical significance of a logical consequence $\Gamma \models \Phi$ is as showing that the following combination of attitudes is (logically) incoherent accepting each of the premises (the sentences in Γ) and rejecting the conclusion (Φ) In a situation where one rejects *multiple* sentences, then it is natural to look slightly more general consequence relation $\Gamma \models \Delta$ is to show that one cannot coherently accept everything in Γ and reject everything in Δ [17] The characterization of multi-conclusion consequence in the classical case is just a straightforward generalization of the usual definition $\Gamma \models \Delta$ holds if and only if every model that makes everything in Γ true makes something in Δ true

---

[16] Some caveats (1) This advantage does not extend to the *disjunction* of all of the individual conjunctions, which the supervaluationist takes to be supertrue, yet is intuitively repugnant (2) Some resist the identification of truth with supertruth see Vann McGee and Brian McLaughlin, "Distinctions without a Difference," *Southern Journal of Philosophy*, Supp Vol xxxii (1994) 203–51 If so, one would need to invoke the principle that one should reject p if and only if p fails to be *determinately* true to get this result Notice, though, that if lack of determinate truth norms rejection in this way, the rest of my argument against local consequence goes through Since these theorists tend to like local consequence, I suggest that they had better give up this account of rejection if so, the supervaluationist who identifies truth with supertruth can plausibly claim an advantage, in that they give a theory which makes the intuitive reaction to the conjunctions above—namely, rejection—the right one

In this "multiple conclusion" setting, one can enquire directly about the coherence of the pattern of acceptance and rejection that was motivated above (recall it is a *good-making* feature of supervaluationism that it is able to preserve this highly intuitive pattern) This is to ask whether the following holds

> A is red, Z is not red $\models_{sv}$
> (A is red $\wedge$ A_2 is not red), (A_2 is red $\wedge$ A_3 is not red), , (A_{n-1} is red $\wedge$ Z is not red)

If $\models$ is fully classical then the answer is straightforward the argument is just as valid as the traditional formulation of the sorites with multiple negated-conjunction premises That this is true follows from a result that is easy to verify that in a classical setting, $\Gamma, \neg\Phi \models \Delta$ iff $\Gamma \models \Delta, \Phi$ [18]

But it would be a disaster if the *supervaluationist* agreed on this score For that would be to grant that all the premises of a sorites argument are acceptable, all the conclusions rejectable, and the argument is valid! The vital connection between questions of validity and questions of the coherence of attitudes would be broken

This directly motivates accepting global rather than local consequence as the correct formulation of supervaluational consequence For global consequence classes the argument invalid on the intended model, the premises are all (super)true, and none of the conclusions are (super)true So the intended model itself provides a countermodel to the putative consequence However, local consequence bids us look at what happens *at each delineation* And whenever the premises are all true at a delineation, at least one of the conclusions will be true So logical consequence agrees with classical consequence that the multi-conclusion formulation of the sorites argument is valid an absurd result from the supervaluationist point of view[19] There are two main ways the defender of local consequence might respond The first is to reject the coherence of multi-conclusion logic[20] The second is to question the *significance* of multi-conclusion logic presupposed above, whereby a valid argument shows us that one cannot simultaneously accept all its premises and reject all its conclusions[21] Such moves seem unnatural to me, especially when there's a straightforward alternative available that makes the problem go away global consequence[22]

---

[17] For an articulation and defense of multiple conclusion systems, see Gregg Restall, "Multiple Conclusions," in P Hajek, L Valdes-Villanueva, and D Westerstahl, eds , *Logic, Methodology and Philosophy of Science Proceedings of the Twelfth International Congress* (London, UK King's College, 2005), pp 189–205, Dominick Hyde, "From Heaps and Gaps to Heaps of Gluts," *Mind*, CVI (1997) 641–60, Brian Weatherson, "Many Many Problems," *Philosophical Quarterly*, LIII (2003) 481–501, and Varzi, *op cit* Restall argues *inter alia* for their utility in analyzing supervaluational "denial " Hyde discusses the departures from classical logic induced by global consequence in a multiple conclusion setting Weatherson appeals to the fact that local consequence induces no revision in multiple conclusion settings as a good-making feature of that account Varzi gives a detailed survey of logical options available to the supervaluationist, including their standing in the relevant multiple conclusion settings

[18] See Restall, *op cit*, p 14

IV THE *PRIMA FACIE* CASE AGAINST REVISIONISM

Global consequence diverges from classical logic when we consider arguments with multiple conclusions and this is a *very good thing* But does $\models_{\text{global}}$ induce logical revisionism even when only a single conclusion is involved? This would be the seemingly "worrying" kind of revisionism that Keefe and Williamson focus upon, giving counterinstances to such entrenched inferential moves as *reductio* and *conditional proof* In this section, I will show that *under a certain assumption* (single-consequence) supervaluational consequence is nonrevisionary In the following sections I will critically examine the assumption

Consider the alleged result that $p \models_{sv} Dp$ When $\models_{sv}$ is read as $\models_{\text{global}}$, we can find counterinstances The very simplest case (there are many others) is the following let the set of sharpenings contains a single delineation δ, relative to which p is true (thus p is supertrue at that model). Assume further that the accessibility relation S relates δ to a delineation δ', and p is false relative to δ'. Then 'Dp' is not true at δ, and hence not supertrue-at-m. Thus we have a countermodel to the claim that Dp is a consequence of p. The upshot is that this sequent cannot play a role in showing that contraposition, or conditional proof, fails.

---

[19] This point can be made without appeal to multiple conclusion logics for the *dual* of supervaluationism subvaluationism (Hyde, *op cit*) The subvaluationists definition of Φ being truth-on-a-model just requires that Φ be true on *some* sharpening This means that all the premises of a standard "long" sorites series will be subtrue (for example, the combination of (1) and (2) and all the negated conjunctions) The standard subvaluationist line is that the "long" sorites fails to be *valid* But this is only true if they work with a globally characterized consequence relation for the argument is *locally valid* This means the sorites will be a valid argument from true premises, with absurdity as its conclusion This seems *clearly* unacceptable the subvaluationist should stick with global validity

[20] This seems mysterious to me after all, I can *define* multi-conclusion logic and tell you its intended significance What could be incoherent about it?

[21] But there *do* seem to be constraints of logical coherence on patterns of acceptance and rejection, and multi-conclusion logic seems the *obvious* way to capture them At a minimum, if one rejects this framework, one owes some substitute account

[22] Varzi, *op cit*, has suggested a fourth alternative that the defender of local consequence might respond by giving a nonstandard "local" reading of what it is for an argument to be sound But the simple fact that there are valid arguments from true premises to untrue conclusions seems bad enough to me

Countermodels can be found to the other examples, too. For example, consider the alleged result that $p \wedge \neg Dp \models_{sv} \perp$.[23] The model above is a counterinstance to this also, for relative to the model just described, the premise is supertrue, but the conclusion superfalse. Indeed, all the results cited fall to such considerations. In an appendix to this paper, I give a simple argument that shows that (for single-premise consequence) global supervaluational consequence coincides with classical consequence.

Game over? Not quite. An opponent need only concede that *some* relation has been constructed which (in single-conclusion cases) behaves exactly like classical consequence. It is still open to the sceptic to argue that there is something illegitimate about the construction: one which means that it does not deserve the name 'supervaluational consequence'. In particular, in setting out my proposal for supervaluational consequence, I let the meaning of 'D' vary from model to model. In the intended interpretation, to be sure, it behaves in a familiar way. But in some unintended interpretations, the value assigned to 'D' will not correspond to anything like the intuitive notion *definitely*.[24]

It is familiar that when offering a semantic characterization of consequence, some interpretations of the language have to be declared "inadmissible." An interpretation of English on which 'or' is interpreted as meaning *and*, would be an interpretation on which 'A' might be true, while 'A or B' false. Nobody takes such interpretations to cast doubt on the validity of or-introduction. Familiarly, we disregard them by a pair of moves: (1) We restrict attention to admissible interpretations of the language: we say that conclusion follows from premises only when all admissible interpretations of the language which make the premises true, make the conclusion true also. (2) We declare that interpretations that do not assign 'and' its intended meaning are inadmissible.

---

[23] Williamson, *op. cit.*, p. 297, footnote 5.19, cites this as a robust result, one from which counterexamples to the above modes of inference all follow.

[24] That observation alone is not enough to cause concern. Few object, for example, when we show the invalidity of $\models$ *Hesperus* = *Phosophorus* by appealing to a model where 'Hesperus' and 'Phosphorus' get assigned distinct referents, and *a fortiori*, one at least is assigned something other than its intended referent.

When characterizing the consequence relation $\models_{\text{global}}$, I presupposed that 'D' is not in the same boat as 'or' in this regard. I took it that I did not need to pause to argue that the interpretations I appealed to did not vary the meaning of 'D'. But one could take issue with this, and insist that interpretations that do not assign 'D' its intended meaning are inadmissible. This may then change the extension of the consequence relation: let us call the new relation $\models^{+}_{\text{global}}$.

For all I have said so far, $\models^{+}_{\text{global}}$ may vindicate all the inferences familiarly imputed to global supervaluational consequence, including in particular the move from p to Dp and the move from $p \wedge \neg Dp$ to absurdity. If so, then by the familiar moves, $\models^{+}_{\text{global}}$ fails to sustain conditional proof, reduction, and the rest. Supervaluational consequence, understood as $\models^{+}_{\text{global}}$, may yet be revisionary.

Two questions therefore arise: first, does $\models^{+}_{\text{global}}$ reinstitute the failures of classical inferential patterns that orthodoxy ascribes to supervaluational consequence? And second, supposing this is so, is $\models^{+}_{\text{global}}$ a better candidate than $\models_{\text{global}}$ for capturing supervaluational consequence?

V. REVISIONISM AND $\models^{+}_{\text{GLOBAL}}$: A VEXED ISSUE

We are concerned in this section with the formal behavior of $\models^{+}_{\text{global}}$, and specifically, whether it leads to failures of conditional proof, reductio, and the rest.

Here is an argument that it does. First, assume that the logic for 'Definitely' is S5, as urged by Keefe.[25] Then, on the intended interpretation, every sharpening S-accesses all and only other sharpenings. Plausibly, the relationship just described between S-accessibility and sharpenings must hold on *every* model that does not shift the meaning of 'Definitely', and only such models are taken into consideration in defining $\models^{+}_{\text{global}}$. Given this, it is easy to see that $p \models^{+}_{\text{global}} Dp$.[26] So, under our assumption, $\models^{+}_{\text{global}}$ will lead to failures of conditional proof, for we also have $\not\models^{+}_{\text{global}} p \supset Dp$.[27]

This discussion proceeded on the assumption that the logic of D was S5. The more popular view, at least in the context of supervaluational treatments of vagueness, is to take it that 'Definitely' obeys some weaker logic that it can be indefinite whether or not *Dp* holds, for example[28] I contend, however, that once we have only these weaker constraints on the accessibility relation, the usual arguments for revisionism lapse, even if we are holding the meaning of 'Definitely' invariant

---

[25] Keefe, *op. cit.*, chapter 7.

[26] *Proof*: Take any model that makes-true p. Then p holds on every sharpening. But that means that p holds on all delineations which any sharpening can S-access, and that is just to day that Dp holds at every sharpening and so is true on the model.

[27] *Proof*: Consider any model where p holds at some sharpenings but not at others. Then $\neg Dp$ holds at all sharpenings. So $p \supset \neg Dp$ fails at those sharpenings where p holds. So the conditional is not (super)true on the model.

To see this, consider the putative validity that Williamson takes to provide the strongest case for revisionism the entailment of absurdity by '$p \wedge \neg Dp$' I will construct a toy model for a vague language which provides a countermodel to the claim that $p \wedge \neg Dp \models^{+}_{\text{global}}$ I hope my reader will agree that nothing in the model is unfaithful to the intended sense of 'Definitely' If so, then this countermodel cannot be regarded as inadmissible The construction turns crucially on an aspect of the supervaluational framework advocated in this paper not shared by Williamson's model. I take it that the difference reflects favorably on the present approach

Suppose that *A*, *B*, *C*, *D*, and *E* are color patches, with each successively redder than the preceding patches Suppose that δ_{AB}, δ_{BC}, δ_{CD} and δ_{DE} are four delineations of a language containing the predicate 'red', corresponding to four placements of the red/non-red cut off in this series (thus δ_{AB} places the cut off between *A* and *B*, and so on) Let us further suppose that the meaning-fixing facts are such that δ_{AB} and δ_{DE} are unintended δ_{BC} and δ_{CD} are the two sharpenings of this interpreted language (see Figure 1) That means that '*A* is red' and '*B* is red' will each be supertrue on this model, since both sharpenings place the cut-off for red no earlier than *B*

Let us further suppose that each cut-off thinks that itself and its neighbors are sharpenings of the language Thus δ_{BC} *S*-accesses itself, δ_{AB} and δ_{CD} (but not δ_{DE}) δ_{CD} *S*-accesses itself, δ_{BC} and δ_{DE} (but not δ_{AB}) (see Figure 2)

On such a model, *A* and *B* will be red, while *D* and *E* will be not-red On the model, it will be indefinite whether *C* is red[29] But in addition, we have "higher order vagueness " On the model, *A* is not only red, it is definitely red, and *E* not only fails to be red, it is definitely not red However, while *B* is red, it is not the case that the model counts it as

---

[28] The coherence of such a statement requires a logic for 'Definitely' weaker than S4 See Williamson, *op cit*, chapter 5, *passim*, for the claim that the logic for definitely should not sustain the characteristic S4 or S5 principles

[29] For this to be the case, it has to be that 'it is indefinite whether *C* is red' is true at each sharpenings, for example, δ_{BC} and δ_{CD}. This requires that each see delineations where *C* is red, and delineations where *C* is not red Since each sees itself and the other, and since they classify *C* opposingly, this condition is met

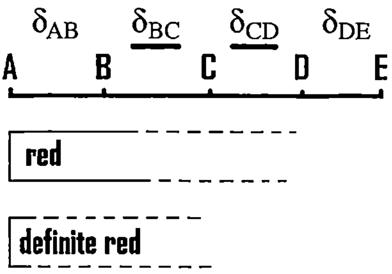


Figure 1

definitely red the sharpening δ_{BC} accesses a delineation δ_{AB} at which it does not count as red[30]

Consider, then the sentence '*B* is red and *D* is not red' This is false at δ_{AB} and δ_{DE}, and is true at δ_{BC} and δ_{CD}. Further, 'It is determinate that *B* is red and *D* is not red' is false not only at δ_{AB} and δ_{DE}, but also at δ_{BC} and δ_{CD}, for each of them sees a nonsharpening delineation (a different one in each case) that makes the sentence false With 'p' = '*B* is red and *D* is not red', it follows that $p \wedge \neg Dp$ is true at this model

I need not argue that this is the *intended* model for some language containing 'Definitely' All that is needed here is the concession that there is no *a priori* reason for ruling out this kind of situation, given the meaning of 'Definitely' For once we admit that $p \wedge \neg Dp$ can be true in some model that preserves the meaning of 'Definitely', then the claim that sentences of that form entail absurdity must be given up, and with it, the argument that $p \wedge \neg Dp \models^{+}_{\text{global}}$ leads to failures of conditional proof, reductio, and the rest[31]

---

[30] Care is needed! It is not the case that the model counts it as definitely red, but neither does it count it as an indefinite case, for *one* of the sharpenings (δ_{CD}) counts it as definitely red

[31] Further, once one appreciates the aspect of the model that lets $p \wedge \neg Dp$ be true on this model—the fact that each sharpenings accesses at least one nonsharpening delineation, one can start to construct models that are less unrealistic For example, one unrealistic aspect of the above model is that it contains only four delineations, and two sharpenings But we can make the same case with a model containing infinitely many of both For each real number r, let δ_r be a delineation that places the cut-offs of 'red' at a wavelength of $r$nm (of light with wavelength less than $r$nm will not count as red at δ_r) There will be a set of admissible cut-offs, corresponding to a range of sharpenings suppose the sharpenings are $\{\delta_r \;\; r_1 > r > r_2\}$ Let $|r_1 - r_2| = d$ Let us define the accessibility relation *S*-access as follows δ_x accesses δ_y iff $|x - y| \leq d$ It is easy to check that this is a reflexive, symmetric but intransitive accessibility relation, and that every sharpening will *S*-access a nonsharpening This is just a generalization of the simpler model given in the text. For example, light of wavelength r_1 will be red, since it counts

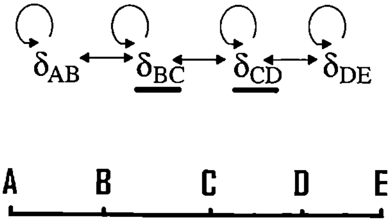


Figure 2

I contend, therefore, that Williamson's route from supervaluationism to logical revisionism is blocked, *even if* we treat 'Definitely' as a logical constant [32]

Why not then simply argue that $\models^{+}_{\text{global}}$, like $\models_{\text{global}}$ is nonrevisionary? I have three reasons for continuing my examination

(1) As mentioned above, there is still a case for revisionism if the logic of 'Definitely' is S5 And that position has some advocates in the literature most notably, Keefe

(2) The case for the nonrevisionism of $\models^{+}_{\text{global}}$ is perhaps not conclusive Perhaps some inventive defender of the orthodox position could make a case that the toy model constructed above, and all others like it, are unfaithful to the intended sense of 'Definitely' (I have no idea what such a case would look like, but cannot rule it out)

---

as red on each sharpening Light of wavelength r_2 will count as nonred, since it is nonred on each sharpening 'Light of wavelength r_1 is red and light of wavelength r_2 is nonred' will be true But since every sharpening sees some (nonsharpening) precisification where this is false, 'Determinately, light of wavelength r_1 is red and light of wavelength r_2 is nonred' will be false at every sharpening, and so the negation is supertrue So again, we have a model making true an instance of $p \wedge \neg Dp$

[32] The critical difference between the characterization of $\models^{+}_{\text{global}}$ given here and the kind of characterization of $\models_{sv}$ that Williamson offers, is the role given to the *sharpenings* In Williamson's setup, supervaluational models do not contain a set of sharpenings rather, they contain a designated "base point." The sharpenings are then *defined* as the set of delineations accessible from the base point That way of formulating matters therefore *guarantees by definition* that there is at least one sharpening which sees all and only other sharpenings (though in general, sharpenings other than the base point may see nonsharpenings) By contrast, if one looks back to the models sketched above, we see that no choice of "base point" will deliver δ_{BC} and δ_{CD} or $\{\delta_r \; r_1 > r > r_2\}$, as the sharpenings

The "base point" formulation Williamson gives, then, is no innocent part of his setup Furthermore, it is an entirely artefactual one the *only* use for a base point is in defining the set of sharpenings, it is the latter notion that plays a crucial role in characterizing truth, consequence, and the like The setup given here, with sharpenings taken as basic rather than artificially defined, is the more natural

(3) All that has been shown here is that one argument (albeit the most famous and influential one) for revisionism fails Unlike the case of $\models_{\text{global}}$, I have offered no *positive* argument that $\models^{+}_{\text{global}}$ is nonrevisionary

For these reasons, further discussion is warranted, and *for the sake of argument* we can assume that $\models^{+}_{\text{global}}$ is revisionary

VI DEFINITELY AND LOGICALITY

Terms which are required to keep their intended meanings across reinterpretations in characterizing logical consequence, are known as *logical constants* $\models^{+}_{\text{global}}$ is defined by insisting that admissible models always feature 'Definitely' with its intended meaning, for $\models_{\text{global}}$, this constraint is lifted We can thus summarize the issue as follows in characterizing supervaluational consequence as $\models_{\text{global}}$, I supposed that D was not a logical constant One who characterizes supervaluational consequence as $\models^{+}_{\text{global}}$, supposes that D is a logical constant

The case for supervaluational consequence sustaining counterexamples to conditional proof, reductio, and the rest, thus at best turns on a highly controversial issue in the philosophy of logic the nature of the logical constraints

Some terms are *paradigmatically* logical 'and', 'or', 'all' are traditional examples It is plausible that it is a constraint on a successful theory of logicality that it categorize these as logical But, moving beyond this traditional set, it is not clear that any informed consensus emerges One looks in vain in the literature on the logic of vagueness for explicit discussion of the putative statue of 'Definitely' as a logical constraint And as with modal and tense operators, I do not see a pre-theoretic case for treating 'Definitely' as a *logical* constant

One might, indeed, turn the considerations given above around, to construct a *prima facie* case against the logicality of 'Definitely' The logical constants, after all, are those terms whose meanings we keep constant in order to generate the logical consequence relation And while we might have little pretheoretic grip on which terms should count as logical constants when we move beyond propositional and predicate logic, the entrenchment of this or that pattern in our inferential practice surely gives us traction on which inferences we are pretheoretically treating as valid It is a presupposition of the debate thus far that conditional proof, reduction, and the rest are entrenched in inferential practice Suppose now it turns out that treating some expression C as a logical constant would mean admitting counterexamples to those rules Absent further considerations, the appropriate response is to take it that we are (implicitly) treating C as nonlogical

At this point in the dialectic, this is exactly the situation that faces us with 'Definitely' [33] Our pretheoretic grip on the goodness of inferential patterns such as *reductio* creates the *prima facie* case that 'Definitely' is not a logical constant This situation should not be overstated if we had strong theoretical reason to treat 'Definitely' as a logical constant, the above considerations will not put up a great deal of resistance But I contend it does shift the burden of proof on the defender of $\models^{+}_{\text{global}}$ to justify their commitments The defender of $\models_{\text{global}}$ wins by default

Whether or not the above is accepted, it is appropriate to consider whether there *is* any theoretical reason to treat 'Definitely' as a logical constant Unfortunately, surveying the literature on logical constants will not help us much, for every consideration we might appeal to is hotly contested

Among those who are happy with talk of a unique set of logical constants, it is controversial what sort of criteria mark off logical from non-logical constants [34] There are those who seek to defend and extend Tarski's permutation criterion for logical constanthood There are those who favor instead inferential criteria logical constants being terms who meaning is exhausted by a specification of introduction and elimination rules There are those who see the demarcation as *pragmatically* drawn as the *minimal* set of constants that are adequate for certain theoretical purposes for the deductive systematization of science, for example And the status of 'Definitely' may well be an issue on which different demarcation criteria return different verdicts [35]

Further, it is controversial, even amongst those who buy into one this or that criterion, how to apply that criteria to "hard" cases (really, anything beyond the "traditional" logical constants of predicate logic) Within the permutation criterion tradition, for example, MacFarlane formulates an account of logicality that is applicable to rich languages such as ones featuring modal and tense operators or the supervaluational setting we are presently working with But the question of whether operators such as 'Always', 'Necessarily', 'Actually', and 'Definitely' count as logical constants turns out to be underdetermined by the permutation criterion alone This illustrates just how far one can push an account of logicality, without have a definitive answer to the "hard cases "[36]

---

[33] I am, of course, supposing something that was left open in the previous section that treating D as a logical constant does lead to counterexamples to conditional proof, and the rest But of course, if the reasoning above is accepted, then we can argue disjunctively either treating D as a logical constant leads to no revision, or if it does, we have a *prima facie* case against treating it as a logical constant. Either way, absent further argument, supervaluational consequence will be nonrevisionary I am also supposing something that will be later called into question that *reductio* and the like are indeed entrenched in inferential practice The point made here is essentially dialectical if it is the case that whether or not supervaluational consequence is damagingly revisionary turns on whether or not we treat 'Definitely' as a logical constant, then that itself creates a *prima facie* case for the conclusion that it should not be so treated Thanks to Ross Cameron for discussion of these issues

[34] I am indebted to John MacFarlane's work on this topic, in particular, *What Does It Mean to Say That Logic Is Formal?* (University of Pittsburgh, PhD Dissertation, 2000), and "Logical Constants," in Edward N Zalta, ed , *The Stanford Encyclopedia of Philosophy* http //plato stanford edu/archives/win2005/entries/logical-constants/ See particularly the latter for an overview of the debate and the labels used below

[35] We can hardly expect, for example, the pragmatic criterion to classify 'Definitely' as logical (in that literature, even the logicality of the first-order existential quantifier is thrown into question)

As well as disagreements about the demarcation principle for logical constants, and disagreements about the applications of such a principle, there are, in addition, there those who would reject the very starting point the idea that there is a privileged set of terms—*the logical constants*—to be theorized about Consider, for example, a view considered explicitly by Tarski

> Perhaps it will be possible to find important objective arguments which will enable us to justify the traditional boundary between logical and extra-logical expressions But I also consider it to be quite possible that investigations will bring no positive results in this direction, so that we shall be compelled to regard such concepts as logical consequence, analytical statement, and tautology as relative concepts which must, on each occasion, be related to a definite, although in greater or less degree arbitrary, division of terms into logical and extra-logical [37]

In the present case, this position would admit *both* $\models^{+}_{\text{global}}$ and $\models_{\text{global}}$ as legitimate consequence relations, but insist that question of which is "really" logical consequence is misguided [38]

There is little hope, therefore, of extracting an answer to the question of whether 'Definitely' is logical, without engaging four-square with this literature I have no space to do this here In light of the *prima facie* case given earlier, I regard $\models_{\text{global}}$ as the pretheoretic front-runner, and if this is sustained, then supervaluational consequence is straightforwardly nonrevisionary However, we cannot ignore the other possibilities canvassed, though not argued for, above that best theory will end up classifying 'Definitely' as a logical constant, or that the very question of whether or not 'Definitely' is a logical constant is misconceived I therefore strengthen my case by arguing in the following section that *even if* 'Definitely' is a logical constant, and even supposing that $\models^{+}_{\text{global}}$ does provide counterexamples to conditional proof and the life, then using the resources constructed in this paper we can show that supervaluational consequence is not *damagingly* revisional in the way often suggested [39]

---

[36] For what it is worth, I believe that MacFarlane's full story will probably classify 'Definitely' and 'Actually' as logical constants But, as he himself emphasizes, there are several places for one sympathetic to his approach to the permutation criterion to resist this contention (see in particular, his discussion of the question of what to count as "intrinsic structure")

[37] Tarski, "On the Concept of Logical Consequence," in *Logic, Semantics, Metamathematics Papers from 1923 to 1938*, J H Woodger, trans (New York Oxford, 1956), pp 402–20, here p 420

[38] I mention this position as illustrative of a range of nondemarcation positions it is the one I find most attractive Another option is provided by "debunkers" such as John Etchemendy (*The Concept of Logical Consequence* (Cambridge Harvard, 1990))

VII NONDAMAGING REVISIONISM

I begin by pointing to an underlying assumption of the debate thus far In stating, for example, that '$\models^{+}_{\text{global}}$ fails to sustain conditional proof', we are assuming that a consequence relation x sustains conditional proof (for example) if and only if the following holds

If A, $B \models_x C$, then $A \models_x B \supset C$

Let us call this mode of inference CP CP is, I take it, the orthodox notion of "conditional proof" within metalogic [40]

However, we should note that the following pattern (which I will call CP*) is sustained

If A, $B \models_{\text{global}} C$, then $A \models^{+}_{\text{global}} B \supset C$

That is, any pattern of reasoning which is valid by the lights of $\models_{\text{global}}$, will make the corresponding conditional a $\models^{+}_{\text{global}}$-validity It would be pointless to start a dispute about whether this pattern should be called 'conditional proof' But the fact CP* holds—whatever we call it—is highly relevant to the overall dialectic To being with, it is relevant to the question of how extensive a revision of classical logic $\models^{+}_{\text{global}}$ induces, assuming turns out to produces counterinstances to CP in the first place In the light of it sustaining CP*, I suggest it would not induce a very extensive revision at all Perhaps "conditional proof" strictly so-called, fails, but every time we move from a $\models_{\text{global}}$-valid argument to the validity of a conditional, we are following a good inference pattern Even if $\models^{+}_{\text{global}}$ turns out to be the one true consequence relation, the only instances of conditional proof that we have to "give up" are those that go from $\models^{+}_{\text{global}}$ valid arguments that are not $\models_{\text{global}}$-valid, to the corresponding conditional

---

[39] I will assume that showing that the relevant candidates for being "logical consequence" are either nonrevisionary, or not damagingly revisionary, will quiet the worries of those who favor nondemarcation views on logicality such as Tarski-inspired relativism

I concentrate in what follows in the (alleged) revisionary implications for *inferential practice* of failures of the kinds of reasoning mentioned above A quite different sort of concern is that *probabilities* (and by extension, degrees of belief) will have to be treated nonclassically if global consequence makes *p*&~*Dp* a contradiction I will not deal with this interesting issue here

[40] Compare McGee and McLaughlin, "Logical Commitment and Semantic Indeterminacy Reply to Williamson," *Linguistics and Philosophy*, XXVII, 1 (2004) 123–36, esp section 3 Though formulated in different terms, the point I make in this section is closely related to their claim that there is an *inferential* version of conditional proof which is available to the supervaluationist even if the full *metatheoretic* one is unavailable

So we have a systematic derived inference pattern—CP* —that just like conditional proof, allows us to draw "categorical conclusions on the basis of hypothetical reasoning "[41] Analogous results for all the other disputed inference patterns [42]

The sense in which the failures of CP or "conditional proof proper" should count as a "revision" must now be reconsidered In a highly theoretical sense, it may be that $\models^{+}_{\text{global}}$ is revisionary, where this is just to express the claim that CP is sustained by the classical logical consequence relation, but not (we are assuming) by $\models^{+}_{\text{global}}$ But I do not think that it was revisionism in this theoretical sense that was supposed to be the *damaging* indictment of supervaluationism Williamson, in particular, talks of the revisionary implications of supervaluational consequence for *inferential practice* (" supervaluations invalidate our natural mode of deductive thinking") [43] But for this charge to be sustained, a case would have to be made that conditional proof proper, and not just CP*, underlies practice That is, we would have to show that inferential practice mandates moving from $\models^{+}_{\text{global}}$-valid but $\models_{\text{global}}$-invalid arguments, to the corresponding conditional No such case has been made, and I contend that none is available If so, then supervaluational consequence is *revisionary of classical logical theory* but not *revisionary of inferential practice* In the light of its conservatism with respect to inferential practice, I contend that whatever revisionary consequences it has will be undamaging

Even if $\models^{+}_{\text{global}}$ rather than $\models_{\text{global}}$ is the correct supervaluational consequence relation, the characterization of $\models_{\text{global}}$ has instrumental value it allows one to formulate rules such as CP* which undermine the Williamson case that supervaluational consequence is revisionary in a damaging sense

---

[41] Williamson, *op cit*, p 152

[42] *Proof* Whenever $\models_{\text{global}} \Phi$, we have $\models^{+}_{\text{global}} \Phi$ Suppose that the $\models_{\text{global}}$ –validity of arguments A_1, A_n entails the $\models_{\text{global}}$-validity of argument C By the above observation, C will be $\models^{+}_{\text{global}}$ –valid also So the rule that takes us from the $\models_{\text{global}}$ –validity of arguments A_1, A_n to the $\models^{+}_{\text{global}}$ –validity of C will hold Reductio, argument by cases and the rest, as well as conditional proof, fit this pattern

[43] Williamson, *op cit*, p 152

VIII. CONCLUSION

One headache for supervaluationists in recent times has been the alleged logical revisionism induced by their semantic framework. I say that there is a natural framework for supervaluationists that undermines the arguments for "damaging" revisionism offered in the literature. Moreover, the setup—involving "extreme" delineations—is one that supervaluationists have independent reasons to accept.

The arguments are undermined in one of two ways. The most straightforward case arises if one accepts my assumption that 'Definitely' is no logical constant. If so, supervaluational consequence demonstrably matches classical single-conclusion consequence (and in particular, conditional proof, reduction, and so on, are retained).

For a setting wherein 'Definitely' is treated as a logical constant, the issue is more vexed. No clean arguments either for or against logical revisionism *stricto sensu* are forthcoming.

Questions of logicality are notoriously hard to arbitrate. It is important, therefore, that we see that even if 'Definitely' does turn out to be a logical constant and that counterexamples to classical modes of inference *stricto sensu* arise, that this in no ways threatens the cogency of inferential practice.

Supervaluational consequence is not revisionary. And even if it is, it is not damagingly so.

J. R. G. WILLIAMS

University of Leeds

APPENDIX

One can think of global supervaluational logic, as defined in this paper, as a modal logic where on each model, sentences are true if they are true at *a set of actual worlds* A rather than at a single actual world a.

Call modal models of the second kind *classical models*, and modal models of the former kind *extended models*. $\Gamma \models_c \Phi$ holds when every classical model M such that every formula in Γ is true-on-M, is such that Φ is true-on-M. $\Gamma \models_{sv} \Phi$ holds when every extended model M' such that every formula in Γ is true-on-M', is such that Φ is true-on-M'.

The operator 'Definitely' is to be thought of as a necessity operator whose accessibility relation (in the simplest cases) is an equivalence relation which on the *intended model* relates distinct worlds iff they are both in A. But the accessibility relation governing 'Definitely' may by very different in *unintended interpretations*.

If we can show that this alteration to modal framework does not change the extension of the consequence relation, then we shall have sustained the central claim of this paper: that global supervaluational consequence is nonrevisionary. For any inferential rules that fails in the extended setting would have to also fail in the setting of a classical modal logic. I now argue that the consequence relation defined via extended models does indeed coincide with the consequence relation defined via classical models.

First, notice that for every classical model, there is an extended model that makes true exactly the same formulas. If a is the actual world of a model M, then let $A = \{a\}$ be the set of actual worlds of an extended model M'. Clearly M and M' make true all the same formulae. This immediately gives us the following:

$$\Gamma \models_{sv} \Phi \Rightarrow \Gamma \models_{c} \Phi$$

So we will have shown the two notions equivalent if we establish the following:

$$\Gamma \not\models_{sv} \Phi \Rightarrow \Gamma \not\models_{c} \Phi$$

Suppose that $\Gamma \not\models_{sv} \Phi$ holds. Then there must be some extended model M^* such that every formulae in Γ was true-on-M^* but Φ was not. For each $\Psi \in \Gamma$ to be true-on-M^*, Ψ has to be true at each world in A_{M^*}. For Φ to fail to be true-on-M^*, it has to be false at some world in A_{M^*}: call this world a^*. Now consider the classical model M^{**} which differs from the extended model M^* just by replacing the set of actual worlds A_{M^*} by the single actual world A^*. By construction, every $\Psi \in \Gamma$ is true at a^* and Φ is false at a^*. Consequently, every $\Psi \in \Gamma$ is true-on-M^{**} and Φ is false-on-M^{**}. Thus, M^{**} provides a countermodel to $\Gamma \not\models_{c} \Phi$. QED

COMMENTS AND CRITICISM

UNDERSTANDING THE DIMENSIONS OF REALIZATION*

Carl Gillett has defended what he calls the "dimensioned" view of the realization relation, which he contrasts with the traditional "flat" view of realization.[1] Intuitively, the dimensioned approach characterizes realization in terms of composition whereas the flat approach views realization in terms of occupiers of functional roles. Elsewhere we have argued that the general view of realization and multiple realization that Gillett advances is not able to discharge the theoretical duties of those relations.[2] Here we focus on an internal objection to Gillett's account and then raise some broader reasons to reject it.

According to Gillett, realization is a relation between property instances that obtains in virtue of the causal powers that the realizing property instances contribute to the individual that instantiates the realized property instance. Specifically:

> Property/relation instance(s) F_1–F_n realize an instance of a property G, in an individual s, *if and only if* s has powers that are individuative of an instance of G in virtue of the powers contributed by F_1–F_n to s or s's constituent(s), but not vice versa.[3]

But from this account of realization and a common assumption that Gillett himself endorses we can produce a *reductio ad absurdum*.

---

*The position of the authors' names reflects only their fondness for alphabetical ordering. We are grateful to Mark Couch, Bob Richardson, Alan Sidelle, Gene Witmer, and audiences at the European Philosophy of Science Association and the Southern Society for Philosophy and Psychology for comments on earlier versions of this paper. Tom Polger's work on this project was supported by the Charles P. Taft Research Center at the University of Cincinnati.

[1] Gillett, "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences," this JOURNAL, C, 11 (November 2003): 591–603; Gillett, "Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction," this JOURNAL, CIV, 4 (April 2007): 193–216. See also Gillett, "The Dimensions of Realization: A Critique of the Standard View," *Analysis*, LXIV, 4 (2002): 316–23.

[2] Lawrence Shapiro, *The Mind Incarnate* (Cambridge: MIT, 2004); Shapiro, "Flat and Happy" (unpublished manuscript); Thomas Polger, *Natural Minds* (Cambridge: MIT, 2004); Polger, "Realization and the Metaphysics of Mind," *Australasian Journal of Philosophy*, LXXXV, 2 (2007): 233–59; Polger, "Two Confusions Concerning Multiple Realizability," *Philosophy of Science*, forthcoming.

[3] Gillett, "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences," p. 594. See also "The Dimensions of Realization: A Critique of the Standard View," p. 322, and "Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction," p. 202. In this latest version, Gillett makes a slight revision to the formulation, about which see note 20, below.

0022-362X/08/0504/213–22

We will first state the argument and then examine its premises.

(P1) Everything that is realized is a property instance, and at least one property instance is realized. (Gillett's account)
(P2) Some things that are realized are multiply realized. (assumption)
(P3) No property instances are multiply realized. (trivial)
(C1) Some property instances are multiply realized. (from P1 and P2)
(C2) Not (P1). (*reductio* from P3 and C1)

The argument is plainly valid, so the only question is about its soundness.

On Gillett's account, realization is defined as a relation between property instances. It follows that only property instances can be realized on his view. Premise (P1) says more than what is said by Gillett's definition of realization alone, for it also asserts that some property instances are in fact realized. But because Gillett argues extensively that many special sciences traffic heavily in realized property instances, and as his account of realization would be of little interest unless it applied to some things, attribution of (P1) to him is fair.

Premise (P2) is a widely shared assumption. One of the chief philosophical interests in the realization relation is precisely that it offers an account of how some entities can be multiply realized.[4] Although there is some dispute as to the exact relation between realization and multiple realization, (P2) makes the fairly weak claim that some things that are realized are also multiply realized. If nothing that is realized was multiply realized, then it would be unclear why realization and multiple realization should be associated.

Premise (P3) says that property instances are not multiply realized. This seems self-evident. Instances are not repeatable. If something is not repeatable, then it is not repeatable in different ways. So property instances are simply not the right sorts of things to be multiply realized.[5]

But (C1) follows immediately from (P1) and (P2). Thus, assuming (P3), Gillett's account of realization leads to contradiction.

---

[4] Hilary Putnam, "The Nature of Mental States" and "Philosophy and Our Mental Life," in his *Mind, Language and Reality: Philosophical Papers, Volume 2* (New York: Cambridge, 1975), pp. 429–40 and 291–303; Ned Block and Jerry Fodor, "What Psychological States Are Not," *Philosophical Review*, LXXXI (1972): 159–81; Fodor, "Special Sciences, or the Disunity of Science as a Working Hypothesis," *Synthese*, XXVI (1974): 97–115; Shapiro, "Multiple Realizations," this JOURNAL, XCVII, 12 (December 2000): 635–54, and *The Mind Incarnate*; Polger, *Natural Minds* and "Realization and the Metaphysics of Mind"; Gillett, "The Dimensions of Realization: A Critique of the Standard View" and "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences."

[5] One might consider the possibility of intra-individual multiple realization, such as variations in the realizers of a particular pain throughout its duration (see, for example, Terence Horgan, "Nonreductive Materialism and the Explanatory Autonomy of Psychology," in Steven J. Wagner and Richard Warner, eds., *Naturalism: A Critical Appraisal* (Notre Dame: University Press, 1993), pp. 295–395. Perhaps in this sense instances

I

We have argued that the idea that property instances can be realized leads to a contradiction So how does Gillett find himself attracted to the view? Gillett's account, as we noted earlier, is designed to contrast with the traditional "flat" account of realization Gillett locates examples of the flat view of realization in the work of Jaegwon Kim and Sydney Shoemaker According to Gillett, "flatties" are committed to the following

(I) A property instance *X* realizes a property instance *Y only if X* and *Y* are instantiated in the same individual

(II) A property instance *X* realizes a property instance *Y only if* the causal powers individuative of the instance of *Y* match causal powers contributed by the instance of *X* (and where *X* may contribute powers not individuative of *Y*) [6]

We believe that conditions (I) and (II) are not in fact principles to which Kim and Shoemaker subscribe That should be welcome news given our argument that a contradiction follows from the view that realization is a relation among property instances [7]

Here are the passages on which Gillett relies From Kim and in support of (I) above, Gillett quotes "It is evident that a second-order property and its realizers are at the same level they are properties of the very same objects "[8] And, from Shoemaker in support of (II) above "property *X* realizes property *Y* just in case the conditional powers bestowed by *Y* are a subset of the conditional powers bestowed by *X*."[9]

Neither Kim's remarks nor Shoemaker's clearly commit either of them to (I) and (II) Both (I) and (II) describe a relationship between property *instances* In contrast, Kim's claim about the relata of realiza-

---

could be multiply realized But an account of multiple realization that covers only such cases, and not the more familiar interindividual cases would be incomplete at best Gillett clearly intends to accommodate interindividual multiple realization, and this is what we say is nonsensical if the realized entities are property instances

[6] Gillett, "The Dimensions of Realization A Critique of the Standard View," pp 317–18 See also Gillett, "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences," p 593

[7] Here a clarification is due Much work on realization—including some of our own previous writings about Gillett's approach—has been either deliberately noncommittal or incautious about whether it is properties, property instances, or whatnot that are the realizing and realized entities These should be seen as attempts to be neutral or even conciliatory while pursuing more general theses about the nature of realization

[8] Kim, *Mind in a Physical World An Essay on the Mind-Body Problem and Mental Causation* (Cambridge MIT, 1998) Italics removed

[9] Shoemaker, "Realization and Mental Causation," in Gillett and Barry Loewer, eds , *Physicalism and Its Discontents* (New York Cambridge, 2001), pp 74–98, see p 78

tion is at least ambiguous. He seems to affirm that it is a property and not a property instance that is realized. Admittedly he refers to both the realized and realizing properties as belonging to the same individual, which is suggestive of reference to property instances. Even so, Kim's remarks do not require, as (I) asserts, that it is a property instance that is realized. Consistent with Kim's passage is the possibility that the property *pain* is realized by a physical state of kind *P*, which is to say that a thing that instantiates *pain* also instantiates physical property *P*. In that case realization remains a relationship between properties rather than property instances—a relation between *P* and *pain*. Shoemaker leaves no room for doubt: he is explicit that he takes realization to be a relation between properties.[10] On our view, realization cannot easily be a relation between properties. And as we argued in the previous section, it definitely cannot be a relation between property instances.[11]

What is wrong with conceiving of realization as a relation between properties? Consider the familiar example of the realization of a sensation such as a pain by a brain process. One point of appealing to realization, at least as we inherit the notion from Hilary Putnam, David Lewis, and Jerry Fodor, is to explain how functionally individuated kinds like Turing machine states, economic exchanges, and mental events can be physically acceptable despite their absence from the fundamental ontologies of the physical sciences.[12] The idea was to show that there is a legitimate way of carving nature at joints other than or in addition to those carved by physics and chemistry. Realization is the special relation by which physical objects like brains can have "nonphysical" states, processes, or properties—perhaps those that are computational, economic, or mental. Physical objects can be instances of "nonphysical" kinds if being so is a matter of what physical objects can do rather than what they are made of. Such "functional" kinds are therefore not "nonphysical" in a way that requires the introduction of any new psychical substance.

---

[10] In his *Physical Realization* (New York: Oxford, 2007), Shoemaker sometimes talks about the realization of property instances, but this appears to be a way of talking about the instantiation of realized properties (for example, p. 12). Unlike Gillett, Shoemaker leaves it open that entities other than properties and property instances can be realized. If Shoemaker insists that property instances are among the things that can be realized, then our critique applies to his view as well.

[11] The cautionary "easily" is necessary simply because we concede that it might be possible to invent a theory of properties or of functions that can make sense of realization in terms of properties. These would not be among the standard views.

[12] Putnam (*op. cit.*), Fodor (*op. cit.*), Lewis, "How to Define Theoretical Terms," this JOURNAL, LXVII, 13 (July 9, 1970): 427–46.

Of course to explain the realization relation we will need to talk about the properties of brains and the properties of, say, pains. After all, the properties of brains help explain what brains can do. But that does not entail that realization is a relation between properties themselves. A realizer of a property, state, or process *P* is a physical thing that meets the criteria that serve to identify instances of *P*, that does the *P* functions. Properties do not "do" anything, in this sense.[13] So the realization relation cannot have properties as both of its relata. Realizers of properties should be *real*—they should be concrete physical tokens.

There is a way in which talking about a physical entity *realizing a property* can be misleading, and this may create ontological confusion. The claim that an object *C* realizes *pain* might mean at least two things:

(1) *C* has the features that are distinctive of the property *pain*.
(2) The property *pain* is made of *C*.

We urge that (1), although not intended as a formal analysis of realization, is the correct way to understand talk about the realization of properties.

Option (2) is a mistake, for if *pain* is a property then it cannot be "made of" some physical object. This is perhaps what leads some theorists to think that realizers must themselves be properties or property instances. But this, as we have said, is a mistake. When one wonders how, if at all, a property *P* may be realized, one is certainly not wondering what properties are made of. One may just as well ask what numbers are made of. The answer might tell us about properties, but it will not tell us about realization. Rather, to ask about the realization of a property is to ask whether there are any physical tokens that exhibit the characteristics that are individuative of the kind of property, state, or process of interest, *pain* for example.

In terms of ontological categories, the property of having some other property or properties is a "second order" or "functional" property. Accordingly, when we said above that the point of realization is to understand how functional properties can be real, we should be taken to mean that the point of realization is to understand whether any physical thing can play the role that characterizes a given functional property. Properties themselves do not play functional roles. When functional properties are realized it is because there are things that occupy the roles that their functional profiles characterize.

---

[13] We do not intend to assume any particular view of properties, here. The simple point is that properties do not have, instantiate, or exemplify the relations characteristic of functional kinds, for example, causal relations. Properties do not stand in the relations that the entities that instantiate them stand in. (No doubt there are kinds of functional relations that can be defined to hold among properties. We can nevertheless be confident that those are not the ones that concern Gillett, Kim, and Shoemaker.)

So it seems that Gillett has misunderstood the flat view when he says that its proponents see realization as a relation between property instances. In fact, flatties have typically described realization as a relation between properties. This, we have argued, is nevertheless a mistake. But it is not a claim that is central to the accounts of Kim or Shoemaker. Rather, we think the core of the flat approach is that it takes realization to be an intralevel relation—namely, the relation of occupying or playing a functional role. When thought of in terms of the realization of properties, this entails that realized and realizer properties must be instantiated in the same individual. That claim corresponds to Gillett's (I), but without the supposition that realization itself is a relation between property instances. That is an improvement of sorts. But we have now given independent reasons against an account of properties as the realizing entities.

II

Gillett claims that the dimensioned analysis of realization is superior to the flat analysis because it is only the former that "covers all the cases of realization illuminated by the sciences."[14] In support of this dramatic claim, Gillett cites his discussion of a diamond's hardness, saying that flatties

> must deny that the alignment and bonding of particular carbon atoms realizes the hardness of the diamond. For neither (I) nor (II) is true in this case, since the properties/relations of the carbon atoms are instantiated in different individuals, and contribute distinct causal powers, from the properties of the diamond. The Flat view thus fails to cover a case of realization and it also erroneously classes the vast swathe of cases drawn from the sciences similar to this one.[15]

Among those cases included in this "vast swathe" are, according to Gillett, many from the chemical and biological sciences, including neuroscience and genetics.[16]

From our perspective, claims like that above for the superiority of the dimensioned analysis are doubly problematic. First, Gillett clearly begs the question against the flat view in the passage above. Of course the flat view fails to identify the properties of carbon atoms and their relations as the realizers of hardness because, according to the flat view, instances of these properties and relations are not even candidates to be realizers of hardness. Because the properties and relations of carbon atoms do not themselves have the causal powers that are individuative of *hard*, they cannot (from the flat perspective) be realizers of hardness.

---

[14] Gillett, "The Metaphysics of Realization, Multiple Realizability, and the Special Sciences," p. 603.

[15] Gillett, "The Dimensions of Realization: A Critique of the Standard View," pp. 319–20.

[16] Gillett, "Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction."

But, second, as Gillett seems to recognize in his molecular case studies, the dimensioned account of realization amounts to no more than a specification of the property-relating relation that parallels the object-relating relation of mereological composition (*ibid.*). According to Gillett, if objects O_1–O_n compose an object M, then properties F_1–F_n of O_1–O_n realize property H of M.[17] Repeatedly he illustrates something that we take to be completely familiar: that the property instances of wholes are noncausally determined by the property instances of their parts. For example, Gillett explains that the instantiation of hardness of a diamond is determined by the instances of properties of the atoms that compose the diamond, and that the spatial memory of a rat is determined by the instances of properties of certain molecules in the rat's brain. He stipulates that this variety of noncausal determination among property instances is "realization."

This account raises two questions. First, why should we accept Gillett's semantic reformation? As we urged earlier, realization was initially introduced as a tool for understanding how functional properties can be instantiated in a physical world. It was not intended to provide an explanation for how composite objects come to instantiate the properties that enable them to be realizers. And second, scientists and philosophers have long known that the properties of big things are dependent on the properties of small things. Thus, it is no surprise that philosophers and scientists have availed themselves of concepts like composition, constitution, noncausal determination, and mereological supervenience in order to capture the relationships between wholes and parts. What does Gillett's dimensioned realization add to this collection of relations? He is in the peculiar position of trying to recruit a label with an already established meaning in order to describe various other relations with their own entrenched labels. One is tempted by Alice's famous response to Humpty Dumpty: "'The question is', Alice said, 'whether you *can* make words mean so many different things' "[18]

---

[17] Gillett is clear that he intends realization to be among a family of interdefined compositional relations. He also defines the relation of *comprising* for the powers he takes to be individuative of properties, and the relation *implementing* for the processes at the level of *O*-entities and the level of *M*-entities. See note 20.

More worrisome than these terminological issues, Gillett's dimensioned "realization" relation obscures a useful distinction. As we note, it is uncontroversial that characteristics of big things are frequently dependent on and determined by the characteristics of the little things of which they are made. Just how the little things manage to do this is a matter of great interest for philosophers and scientists alike, and there is no *a priori* reason to suppose that there is only one way. The relationships that mere aggregates have to their components may differ from the relationship that organized wholes have to theirs. Compare how grains of sand contribute to a pile to how a single gas molecule's kinetic energy contributes to the temperature of a volume of gas, to how carbon atoms contribute to the properties of a diamond. And these may differ still from the ways that the constituents of functional entities determine the properties of a mechanism, such as the contributions of a lever and screw to a corkscrew. Gillett's dimensioned account lumps these together and calls them all "realization," and we think this is a mistake.[19]

Notice that Gillett's formulation of realization asserts that the entity in which an instance of property *G* is realized has the powers individuative of *G* "in virtue of" the properties that it or its constituents have. But this runs together all the different ways—all the different dependence and determination relations, all the different "in virtue of" relations—by which the properties of parts may contribute to the property of a whole. Insofar as dimensioned realization gives the appearance of explaining something related to what the "flatties" are trying to explain, it seems to us that other kinds of dependence relations—relations already handy in any metaphysician's toolkit—are in fact doing all the explanatory work. Worse, Gillett replaces these relations with one that is explanatorily empty: to say that the properties of the atoms of the diamond "realize" the properties of the diamond is for Gillett just to reiterate that the latter depend on the former—that they are instantiated "in virtue of" the former. This is a classic example of *virtus dormitiva* explanation: the properties of the atoms determine the properties of the diamond because the diamond is made of the atoms and thereby has its properties in virtue of the properties of the atoms.[20]

---

[18] Lewis Carroll, *Through the Looking-Glass and What Alice Found There* (New York: Macmillan, 1899), p. 123.

[19] One might, of course, *argue* that they can all be reduced to a single relation. We resist this suggestion, but at any rate Gillett has furnished no such argument.

[20] Gillett ("Understanding the New Reductionism: The Metaphysics of Science and Compositional Reduction," p. 202) replaces the "in virtue of" formulation with one

In contrast, our approach to understanding realization aims to add to the metaphysicians toolkit: some properties of wholes are composed, some constituted, some realized, and so on. According to our view, property, state, or process kind *P* is realized when something plays the functional role characteristic of *P*. Of course each of these dependence relations requires further examination. We, for example, owe an account of functional roles and role playing.[21] But it is no good to explicate such relations by reporting that they are dependence relations between entities of certain classes—objects, properties, or whatnot. For that is simply to describe the phenomenon that we aim to explain.

It may be tempting at this point to say that we and Gillett are simply talking about different relations. The "flat" and "dimensioned" approaches are accounts of two different kinds of realization, one could say. This response has recently attracted some adherents.[22] But we think this is a mistake akin to allowing that China's "democracy" is just a different kind of democracy than we have in the West. Gillett says that an account of realization is needed, and we agree. He offers an account. In our assessment, Gillett is wrong. To respond that he is simply talking about a different relation than we are is to sugarcoat the serious problems we have identified with his conception of realization. We see no good reason to say that he is right about something called "realization" if he is wrong about realization.

III

Gillett's dimensioned analysis of realization confronts the following challenges. First, the account is incoherent insofar as it makes multi-

---

according to which F_1–F_n realize *G* if and only if their powers "together comprise the powers individuative of *G*." Comprising occurs, according to Gillett, when the powers C_1–C_n of F_1–F_n "together implement the mechanisms grounded by the triggering and manifestation of" the comprised power C^*. Gillett does not explicate the *implementation* relation between processes or mechanisms in this paper, aside from remarking that "the scientific notions of comprising, constitution, or parthood, and implementation are like realization in the relevant component entities being such that together they noncausally result in the composed entity, but not vice versa" (p. 205). But elsewhere he defines implementation in terms of comprising, realization, and parthood and composition (Gillett, "Making Sense of Levels in the Sciences," unpublished manuscript, see also "Understanding the New Reductionism," fn. 13). It is hard to see that this system avoids our complaint that Gillett is only giving a name to the various dependence relations, rather than explaining them.

[21] In *Natural Minds,* Polger begins to make good on this debt, for various notions of function and role.

[22] For example, Robert Wilson and Carl Craver, "Realization," in Paul Thagard, ed., *Handbook of the Philosophy of Science, Volume 12: Philosophy of Psychology and Cognitive Science* (The Netherlands: Elsevier, 2007), pp. 81–104. See also Ronald Endicott, "Flat versus Dimensioned: The What and the How of Realization" (unpublished manuscript).

Adrian M. S. Piper

Rationality and the Structure of the Self

A Two-Volume Study in Kantian Metaethics

Adrian Piper's practice as a first-generation Conceptual artist in the 1960s led her to Kant's *Critique of Pure Reason,* and this, in turn, to an in-depth study of philosophy that culminated in a Ph.D. in philosophy from Harvard and a successful academic career as a professor of philosophy. Though Piper has referred to this as her "day job," she has also often said that her philosophical work is the theory that her artwork puts into practice. That theory is now available in its entirety in her *Rationality and the Structure of the Self. Volume I: The Humean Conception* critiques the prevailing desire-centered conception of the self, and *Volume II: A Kantian Conception* develops an original Kantian alternative. Encompassing the history of moral and political philosophy as well as engaging head-on with contemporary debates in ethics, economics and philosophical psychology, Piper's two-volume project has been praised by philosophers as "groundbreaking," "very powerful," "original and important," "brilliant," "indispensable," and "a highly significant contribution." Art professionals will find it equally useful for understanding Piper's unique contribution to Conceptual art and to art theory more generally.

RATIONALITY AND THE STRUCTURE OF THE SELF, *Volume I: The Humean Conception*

The Humean conception of the self consists in the belief-desire model of motivation and the utility-maximizing model of rationality. This conception has dominated Western thought in philosophy and the social sciences ever since Hobbes' initial formulation in *Leviathan* and Hume's elaboration in the *Treatise of Human Nature*. Bentham, Freud, Ramsey, Skinner, Allais, von Neumann and Morgenstern and others have added further refinements that have brought it to a high degree of formal sophistication. Late twentieth century moral philosophers such as Rawls, Brandt, Frankfurt, Nagel and Williams have taken it for granted, and have made use of it to supply metaethical foundations for a wide variety of normative moral theories. But the Humean conception of the self also leads to seemingly insoluble problems about moral motivation, rational final ends, and moral justification. Can it be made to work?

RATIONALITY AND THE STRUCTURE OF THE SELF, *Volume II: A Kantian Conception*

Adrian Piper argues that the Humean conception can be made to work only if it is placed in the context of a wider and genuinely universal conception of the self, whose origins are to be found in Kant's *Critique of Pure Reason*. This conception comprises the basic canons of classical logic, which provide both a model of motivation and a model of rationality. They also supply necessary conditions both for the coherence and integrity of the self and also for unified agency. The Kantian conception solves certain intractable problems in decision theory by integrating it into classical predicate logic, and provides answers to longstanding controversies in metaethics concerning moral motivation, the nature of rational final ends, and moral justification that the Humean conception engenders. In addition, it sheds light on certain kinds of moral behavior – for example, the whistleblower – that the Humean conception is at a loss to explain.

Both volumes of *Rationality and the Structure of the Self,* plus seven press readers' reports, are available for free download at http://www.adrianpiper.com/rss/index.shtml

STONY BROOK
STATE UNIVERSITY OF NEW YORK
The Department of Philosophy presents

Stony Brook Manhattan
401 Park Avenue South
Between 27th and 28th Street

COGNITIVE DISABILITY:
A CHALLENGE TO MORAL PHILOSOPHY

Rethinking and revising traditional assumptions about personhood, autonomy, agency, human relationships, justice, care, classification, and moral responsibility in light of Alzheimer's, autism and 'mental retardation'

SEPTEMBER 18 – 20, 2008

Speakers include
Michael Berube, Ian Hacking, Jeff McMahan, Martha Nussbaum, Peter Singer, Daniel Wikler, Jonathan Wolff, Licia Carlson, Leslie Francis, Agnieszka Jaworska, Bruce Jennings, Eva Kittay, Hilde Lindemann, Victoria McGeer, James Nelson, Hans Reinders, Henry Richardson, Anita Silvers, Cynthia Stark, David Shoemaker, Anna Stubblefield, Sophia Wong

For more information
www stonybrook edu/cdconference
Eva Kittay@gmail com

Co-Directors
Eva Kittay Stony Brook University/SUNY
Licia Carlson Harvard University

Co-sponsored by The Center for Discovery, New York University Bioethics Center, *Metaphilosophy* (Wiley-Blackwell), and the Office of the Provost at Stony Brook University/SUNY
The Alzheimers Association, the College of Arts and Sciences, and the Humanities Institute at Stony Brook University/SUNY, Department of Philosophy at Southern Connecticut State University, Stony Brook University's Templeton Research Lecture Series

For disability related accommodation contact Licia Carlson at (617) 496 4349
Stony Brook University/SUNY is an affirmative action equal opportunity educator and employer

You can now subscribe to the *Journal of Philosophy* using

PayPal

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org.**

HE RUTGERS UNIVERSITY PHILOSOPHY DEPARTMENT IS PLEASED TO HOST:

ARISTOTLE ON PREDICATION

RUTGERS UNIVERSITY INN AND CONFERENCE CENTER

DOUGLASS CAMPUS, NEW BRUNSWICK, NJ

OCTOBER 4-5, 2008

two day conference devoted to an examination of the treatment of predication in arious Aristotelian treatises, including the *Categories*, *De Interpretatione*, the *Topics*, e *Analytics*, *De Anima*, and the *Metaphysics*

rganized by David Charles (Oriel College, Oxford University), Alan Code (Rutgers niversity), and Christof Rapp (Humboldt Universitat zu Berlin)

nvited speakers and participants include:

onathan Beere (Humboldt Universitat zu Berlin)
obert Bolton (Rutgers University)
Valter Cavini (University of Bologna)
ei Chiba (Hokkaido University)
aolo Crivelli (New College, Oxford University)
rank Lewis (University of Southern California)
arko Malink (Humboldt Universitat zu Berlin)
ichard Patterson (Emory University)
nnamaria Schiaparelli (Queen's College, Oxford University)
obin Smith (Texas A&M)
iko Strobach (Universitat des Saarlandes, (Saarbrucken))

nyone in the world may attend, but all participants must register by September 12 by mailing Carrie Swanson at nous@eden rutgers edu You will be contacted about onference registration details and lodging Further information will be made available at ttp //aristotle rutgers edu/

CAMBRIDGE

New and Noteworthy Titles

Philosophy from a Skeptical Perspective
Joseph Agassi and Abraham Meidan
$80.00 Hb 978-0-521-89812-6
$22.99 Pb 978-0-521-72639-9
190 pp

Leibniz
An Intellectual Biography
Maria Rosa Antognazza
$39.95 Hb 978-0-521-80619-0
550 pp

Demystifying Legal Reasoning
Larry Alexander and Emily Sherwin
Cambridge Introductions to Philosophy and Law
$85.00 Hb 978-0-521-87898-2
$26.99 Pb 978-0-521-70395-6
264 pp

The Cambridge Dictionary of Bioethics
Edited by Robert B. Baker and Laurence B. McCullough
$250.00* Hb 978-0-521-80709-8
960 pp

The Paradox of Predictivism
Eric Christian Barnes
$99.00 Hb 978-0-521-87962-0
276 pp

The Problem of Punishment
David Boonin
$85.00 Hb 978-0-521-88316-0
$27.99 Pb 978-0-521-70961-3
310 pp

Ethics of Global Development
Agency, Capability, and Deliberative Democracy
David A. Crocker
$99.00 Hb 978-0-521-88519-5
430 pp

French Political Thought from Montesquieu to Tocqueville
Liberty in a Levelled Society?
Annelien de Dijn
Ideas in Context
$99.00 Hb 978-0-521-87788-6
230 pp

What is Analytic Philosophy?
Hans-Johann Glock
$90.00 Hb 978-0-521-87267-6
$24.99 Pb 978-0-521-69426-1
304 pp

An Introduction to Africana Philosophy
Lewis R. Gordon
Cambridge Introductions to Philosophy
$90.00 Hb 978-0-521-85885-4
$29.99 Pb 978-0-521-67546-8
288 pp

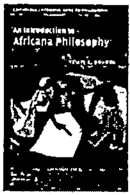
Plato *The Symposium*
Translated by M. C. Howatson
Edited by Frisbee C. C. Sheffield
Cambridge Texts in the History of Philosophy
$49.00 Hb 978-0-521-86440-4
$13.99 Pb 978-0-521-68298-5
128 pp

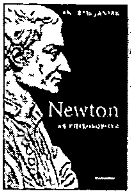
The Cambridge Companion to Galen
Edited by R. J. Hankinson
Cambridge Companions to Philosophy
$85.00 Hb 978-0-521-81954-1
$29.99 Pb 978-0-521-52558-9
472 pp

Newton as Philosopher
Andrew Janiak
$90.00 Hb 978-0-521-86286-8
208 pp

Hegel's *Phenomenology of Spirit*
An Introduction
Larry Krasnoff
Cambridge Introductions to Key Philosophical Texts
$80.00 Hb 978-0-521-87357-4
$27.99 Pb 978-0-521-69537-4
184 pp

Prices subject to change





www.cambridge.org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

from Cambridge

An Introduction to Chinese Philosophy
Karyn L Lai
Cambridge Introductions to Philosophy
$90 00 Hb 978 0-521-84646-2
$34 99 Pb 978 0-521-60892-3
328 pp

The Autonomy of Morality
Charles Larmore
$85 00 Hb 978 0 521 88913-1
$24 99 Pb 978-0 521-71782-3
290 pp

Aggression and Crimes Against Peace
Larry May
$85 00 Hb 978-0 521-89431-9
$27 99 Pb 978-0-521-71915-5
368 pp

War
Essays in Political Philosophy
Edited by Larry May
$85 00 Hb 978-0 521-87637-7
$28 99 Pb 978-0-521-70004-7
344 pp

The Elusive God
Reorienting Religious Epistemology
Paul K Moser
$90 00 Hb 978 0-521-88903-2
304 pp

Hegel's *Phenomenology of Spirit*
A Critical Guide
Edited by Dean Moyar and Michael Quante
Cambridge Critical Guides
$99 00 Hb 978-0-521-87454 0
258 pp

An Introduction to the Philosophy of Religion
Michael J Murray and Michael C. Rea
Cambridge Introductions to Philosophy
$90 00 Hb 978-0 521-85369 9
$29 99 Pb 978-0-521 61955 4
306 pp

Toward a Theory of Human Rights
Religion, Law, Courts
Michael J Perry
$29 99 Pb 978-0-521-68422 4
253 pp

An Introduction to Non-Classical Logic
2nd Edition
From If to Is
Graham Priest
Cambridge Introductions to Philosophy
$99 00 Hb 978-0-521-85433 7
$34 99 Pb 978 0-521-67026 5
648 pp

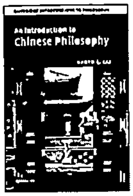
Evidence and Evolution
The Logic Behind the Science
Elliott Sober
$90 00 Hb 978 0-521-87188-4
$29 99 Pb 978 0-521 69274-8
412 pp

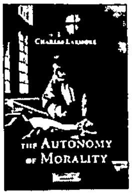
Phenomenology of the Human Person
Robert Sokolowski
$85 00 Hb 978-0-521-88891-2
$26 99 Pb 978 0 521-71766 3
360 pp

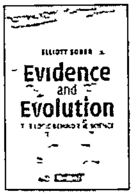
Information Technology and Moral Philosophy
Edited by Jeroen van den Hoven and John Weckert
Cambridge Studies in Philosophy and Public Policy
$90 00 Hb 978-0-521-85549-5
432 pp

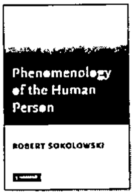
Argumentation Schemes
Douglas Walton, Christopher Reed, and Fabrizio Macagno
$90 00 Hb 978 0 521 89790-7
$32 99 Pb 978-0-521-72374 9
436 pp

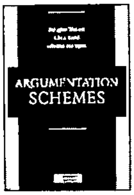
Prices subject to change











www cambridge org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS

New from OXFORD

NEW IN PAPERBACK

CHOOSING CHILDREN

GENES, DISABILITY, AND DESIGN

JONATHAN GLOVER

"Delightfully readable, insightful essays Glover writes with enormous authority a thoughtful and superbly written book"–*BBC Focus*

(Uehiro Series in Practical Ethics)

2006 (paper 2008) 128 pp

paper $13 99 cloth $17 95

DISJUNCTIVISM

PERCEPTION, ACTION, KNOWLEDGE

Edited by ADRIAN HADDOCK and FIONA MACPHERSON

Disjunctivism is the focus of a lively debate spanning the philosophy of perception, epistemology, and the philosophy of action Adrian Haddock and Fiona Macpherson present seventeen specially written essays, which examine the different forms of disjunctivism and explore the connections between them

2008 400 pp $110 00

PLATO: LAWS 10

TRANSLATED WITH AN INTRODUCTION AND COMMENTARY

ROBERT MAYHEW

Book 10 of the *Laws* sets out Plato's last thoughts on the nature and existence of gods and souls, the role of the gods in the cosmos, the connection between piety and virtue, and the role of religion in a properly organized city Robert Mayhew presents a new English translation of this important text with a detailed commentary that highlights its philosophical, political, and religious significance

(Clarendon Plato Series)

2008 240 pp $70 00

RATIONALITY AND COMMITMENT

Edited by FABIENNE PETER and HANS BERNHARD SCHMID

In *Rationality and Commitment*, thirteen leading philosophers and economists discuss the Nobel Laureate Amartya Sen's trenchant critique of rational choice theory, and propose their own answers to the question of how to account for the rationality of committed action

2008 314 pp $80 00

PLEASURE, MIND, AND SOUL

SELECTED PAPERS IN ANCIENT PHILOSOPHY

C C W TAYLOR

C C W Taylor presents a selection of his writings on a set of important topics in the work of some of the greatest ancient philosophers The central theme of the volume is the moral psychology of Plato and Aristotle, with a speical focus on pleasure and related concepts, an area central to Greek ethical thought

2008 320 pp $90 00

THE ETHICS OF PHILODEMUS

VOULA TSOUNA

Tsouna examines Philodemus's theoretical principles in ethics, his contributions to moral psychology, his method, his conception of therapy, and his therapeutic techniques *The Ethics of Philodemus* will be of considerable interest to scholars and students of classics, ancient philosophy, and ethics

2008 280 pp $72 00

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US To order, please contact customer service at 1-866-445 8685, or visit us online at www oup com/us

12.1.09
UNIVERSITY
OF
CALCUTTA

THE JOURNAL OF PHILOSOPHY

VOLUME CV, NUMBER 5
MAY 2008

page
223 *Counterfactual Intervention and Agents' Capacities* — Neil Levy
240 *On Having a Point of View Belief, Action, and Egocentric States* — Cheryl K. Chen

COMMENTS AND CRITICISM
259 *Deontology, Individualism, and Uncertainty A Reply to Jackson and Smith* — Ron Aboodi, Adi Borer, and David Enoch
273 *Moral Absolutism Defended* — Patrick Hawley
276 IN MEMORIAM MARY MOTHERSILL

Published by **The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Michael J Mooney, *Treasurer,* Liaquat Ahamed, *Other Trustees* Lee Bollinger, Leigh S Cauman, Barbara Gimbel, Kent Greenawalt, Lynn Nesbit, Daniel Shapiro

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at www journalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2008

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER. Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2008 by the Journal of Philosophy, Inc ISSN 0022–362X

CENTRAL LIBRARY
UNIVERSITY OF CALCUTTA

THE JOURNAL OF PHILOSOPHY
VOLUME CV, NO. 5, MAY 2008

COUNTERFACTUAL INTERVENTION AND AGENTS' CAPACITIES*

It is widely held that Frankfurt-style cases (FSCs) decisively alter the landscape of debates over free will.[1] Prior to their emergence, the debate between compatibilists and incompatibilists seemed irretrievably bogged down in the morass created by differing intuitions and different analyses of the central terms upon which the debate turned. For each of the central concepts—*abilities, alternative possibilities, capacities*—there were compatibilist and incompatibilist analyses available, and there seemed to be no decisive, nonquestion-begging, reasons to prefer one kind to the other. What John Martin Fischer has usefully dubbed a *dialectical stalemate* ensued.[2] FSCs changed all that, by showing (to the satisfaction of most) that alternative possibilities, and therefore agents' abilities to actualize such possibilities, are not necessary for moral responsibility. FSCs demonstrated that moral responsibility depends only upon *actual sequence* properties of agents, not on what might happen in various counterfactual scenarios. Though agents lack alternative possibilities in FSCs, the features that explain why they lack such alternatives play no role in the explanation of what they do. Given that fact, agents in FSC act "on their own," and are therefore apparently responsible for what they do.

---

* I would like to thank John Martin Fischer and audiences at Oxford University, especially Joe Raz, the Australian National University, especially David Chalmers, and Bristol University for helpful comments on this paper. I also owe a debt of gratitude to three anonymous referees for this JOURNAL.

[1] Frankfurt-style cases are so called because they were first introduced—to the modern debate, at least—by Harry Frankfurt, in "Alternate Possibilities and Moral Responsibility," this JOURNAL, LXVI, 23 (December 4, 1969): 829–39.

[2] Fischer, *The Metaphysics of Free Will* (Cambridge: Blackwell, 1994), pp. 83f.

0022-362X/08/0505/223–39

In this paper, I shall argue that this conclusion is unwarranted FSCs can show that agents are responsible for their actions, despite lacking alternative possibilities, only if we are justified in believing not only that the agents they feature act "on their own," but also that their actions are an expression of *their* normal capacities for evaluating and responding to reasons And we are justified in thinking that agents in FSCs act in a way that expresses their normal capacities only if we are justified in distinguishing between the intrinsic, responsibility-ensuring, capacities of the agent, and features of her context I shall show that there is no principled, non-*ad hoc*, manner of making this latter distinction I shall do this fighting thought experiment with thought experiment by introducing a new kind of case, Frankfurt-enabler cases (FECs), in which agents apparently *gain* a capacity due to the presence of a counterfactual intervener In FECs, the intuition that agents gain a capacity due to the presence of a counterfactual intervener depends upon our regarding the agent as an ensemble agent-plus-intervener But there is no principled reason, I claim, why we should distinguish between the agent and her context in FSCs, but fail to make this distinction in FECs We should treat the cases alike I argue that we can never understand agents' capacities, and thus the sensitivity to reasons of the mechanisms upon which they act, except by considering the context in which they are embedded Hence, we are not entitled to conclude that the capacities of agents in FSCs are unchanged by the presence of the counterfactual intervener

In the first section of the paper, I shall set out the argument for the view I aim to refute, which holds that agents' capacities are entirely dependent upon their intrinsic properties I shall concentrate here on Fischer and Mark Ravizza's influential account of moral responsibility, which has the virtue (among many others) of explicitly developing this approach to capacities In the following section, again by way of reflection on Fischer and Ravizza, I shall show that the distinction between the agent and her context cannot be consistently maintained when we are assessing her capacities, and that therefore we are not entitled to the conclusion that agents in FSCs retain their normal capacities I shall conclude by considering the relationship between the argument I develop here and some recent work in philosophy of mind on the location of mental states

I FRANKFURT-STYLE CASES AND AGENTS' ABILITIES

In an FSC, an agent performs a morally significant action—say, voting for a political candidate in an election—in what is apparently the normal way, and therefore seems morally responsible for her action However, due to circumstances of which the agent is unaware, she lacks alternatives she cannot do otherwise than choose to vote in the way she does Consider the following, more or less typical, FSC

> *Voting* When she enters the voting booth, Connie had not yet made up her mind whether to vote Democratic or Republican Unbeknownst to her, an evil but gifted neuroscientist monitors her neural states using a computer chip he has implanted in her brain The chip gives the neuroscientist the power to intervene in Connie's neural processes Connie is psychologically constituted so that it is a necessary (though not sufficient) condition of her voting Democratic that she thinks deeply about a certain policy, were Connie to think deeply about this policy, the neuroscientist would intervene, causing her to choose to vote Republican But Connie does not think deeply about the policy and proceeds to vote Republican on her own The neuroscientist and his device play no role in bringing about Connie's vote [3]

Connie lacks any alternative to voting Republican, were she to think deeply about the policy that might motivate her to vote Democratic, the neuroscientist would intervene to bring it about that she votes Republican But since she acts entirely on her own, she seems responsible for deciding to vote Republican, and for voting Republican

On the basis of cases like this, Frankfurt and his followers conclude that moral responsibility does not require alternative possibilities Since the conditions that ensure that Connie votes as she does play no role in her *actual* voting, these ensuring conditions cannot be relevant to her moral responsibility FSCs have played a pivotal role in motivating new kinds of accounts of free will, compatibilist and libertarian

If agents are responsible in FSCs, despite their lack of alternatives, there must be something about the *actual* sequence that grounds their responsibility Accounting for responsibility in such cases, and distinguishing between responsible agents and those subject to the standard excuses and justifications, therefore requires us to (as Fischer puts it) decode the actual sequence [4] FSCs are important because they enable us to see that what might happen in counterfactual scenarios is an irrelevant distraction when it comes to assessing moral responsibility (though counterfactuals remain important as a means of getting at what is happening in the *actual* sequence). FSCs force us to focus on the actual sequence and the properties of the agent in that sequence. By focusing on these properties, we can isolate the characteristics of agents that make them responsible.

---

[3] The FSC sketched here is a variation on the voting case frequently used by Fischer, with an addition borrowed from Derk Pereboom The addition is that the neuroscientist's signal for intervention is a thought that is necessary but not sufficient for the agent's deciding in a manner different to the way the neuroscientist wants If the signal is a sufficient condition for her making the undesired decision, then Connie might be said to have begun the process of deciding to vote Democratic, she would have robust alternatives after all (to vote Republican or to begin to vote Democratic) But if the signal is only a necessary condition, Connie seems stripped of the ability to vote Democratic, yet still seems to vote in the ordinary manner For the original voting case, see Fischer, *My Way* (New York Oxford, 2006), pp 41 and elsewhere, for Pereboom's case, see Pereboom, *Living without Free Will* (New York Cambridge, 2001), pp 18–38

[4] Fischer, *My Way*, p 7

The clearest exposition of the way FSCs work to isolate the relevant characteristics of agents is due to Fischer, alone and together with Ravizza. Fischer argues that agents are responsible for acts and omissions just in case they exercise *guidance control* over their bodily movements and over the relevant states of affairs in the world, where "guidance control" is a kind of counterfactual dependence of actions upon reasons. Agents exercise guidance control when the mechanism upon which they act is moderately reasons-responsive; moderate reasons-responsive consists, in turn, of regular sensitivity to reasons, including moral reasons, and (at least) minimal reactivity to reasons. That is, a mechanism is moderately reasons-responsive if it is capable of tracking reasons in some kind of patterned way, and in addition there is at least one possible world in which it would cause a different action in response to a reason.[5]

Fischer argues that we test whether a mechanism is moderately reasons-responsive by asking how it would respond in various counterfactual scenarios. This test is similar to, but importantly different from, the test applied by traditional compatibilists. Traditional compatibilists held that an agent is able to do otherwise iff (roughly) were she to desire to do otherwise, she *would* do otherwise. The problems with this conditional analysis of ability are well-known: among other things, it analyzes 'ability' in terms that seem themselves to presuppose that very notion (in the form of the ability to desire otherwise). Fischer's test avoids these problems. He does not ask whether an agent would do otherwise in the nearest possible world, or in the nearest world in which she desired to do otherwise, nor does he take the upshot of the conditional to bear upon whether the agent possesses the ability to do otherwise, where 'ability' is understood in the 'all in' sense.[6] Instead, he asks how the agent would respond in various counterfactual scenarios only in order to discover the properties of the mechanism upon which the agent *actually* acts. If it is the case that the mechanism would cause the agent to act otherwise in some possible world in which she was presented with a sufficient reason to do otherwise *and* in which the counterfactual intervener was not present, then the mechanism is properly reactive to reasons, similar tests would tell us whether it was properly responsive. The worlds which are relevant to these counterfactuals need not be accessible to the agent, we are not concerned with whether the agent *can* do otherwise. The lesson of FSCs is that this 'all in' ability is irrelevant to moral responsibility. What matters is what happens in the actual sequence—specifically, what are the properties of the mechanism upon which the agent acts—and the relevant counterfactuals give us this information.

---

[5] Fischer and Ravizza defend their notions of moderate reasons-responsiveness and guidance control in their *Responsibility and Control: An Essay on Moral Responsibility* (New York: Cambridge, 1998). In addition to reasons-responsiveness, a mechanism must satisfy an ownership condition if the agent is to be responsible for actions upon it. The ownership condition will play no role here and I set it aside.

[6] See Fischer, *My Way*, p. 140 n. 22.

We might profitably compare the questions Fischer asks, in order to identify the properties of the mechanism upon which the agent acts, with the kinds of questions many philosophers urge us to ask in order to discover whether an object possess a particular disposition. Just as the simple conditional analysis of 'ability' is refuted by FSCs, so a simple conditional analysis of dispositions is refuted by cases in which the disposition is *finkish*.[7] A disposition is finkish (roughly) iff the conditions for eliciting the disposition are also sufficient to prevent its manifestation. Suppose, for instance, that a powerful sorcerer took a liking to a particular vase. The vase might be extremely fragile, intuitively, it would remain fragile even if the sorcerer took it into his head to protect it in the following manner were the vase to be dropped, the sorcerer would cast a spell that prevented it shattering. Fragility is a property that is constituted by the physical structure of the vase the nature of the bonds between the molecules of which it consists. Thus, the vase possesses the disposition right now, as it sits undisturbed on the shelf, even though it is false that were it to fall it would shatter. The sorcerer is a counterfactual intervener, closely analogous to the neuroscientist of the FSC, and like the latter his presence enables us to see how dispositions and capacities are constituted by intrinsic properties of (respectively) objects and agents. Just as the vase possesses the disposition to shatter, despite the fact that—due to the presence of the sorcerer—it will not shatter if dropped, so the agent in an FSC possesses the kinds of capacities that make her morally responsible, though—due to the presence of the neuroscientist—she cannot exercise these capacities.

The fragility of the vase is constituted by the bonds between its molecules. Fragility is a property the vase possesses right now, while it sits on the shelf, whether or not it would continue to possess that property were the sorcerer to cast his spell (the vase would continue to possess the property in a *masked* way if the spell would prevent the shattering while leaving the molecular structure untouched; the disposition would be finked, and lost, perhaps temporarily, if the spell were to alter the molecular structure[8]). Fragility is, at were, an actual sequence property, which the vase has *now*, regardless of what might happen in nearby worlds. We test for fragility not by asking what would happen in the closest world in which the vase was dropped, but by asking how it would behave in a world in which the vase was dropped *and* the sorcerer did not cast his spell. We ask about that world because so doing is the best way to understand the properties the vase has in *this* world. Similarly, in assessing whether an agent is responsible for a decision or an action, we ask how the mechanism upon which she actually acts would respond in a world in which the agent was presented with sufficient reason to do otherwise *and* the counterfactual intervener failed to act. We ask about that world because this is the only way to understand the capacities of the relevant mechanism in *this* world. In each case, the counterfactuals are relevant not because we are interested in what happens in alternative scenarios, for the sake of such alternatives, but because we are interested in the properties of the actual sequence.

---

[7] C B Martin, "Dispositions and Conditionals," *The Philosophical Quarterly*, XLIV (1994) 1–8

Some philosophers have hoped to analyze 'ability' on the lines suggested by the example of finkish dispositions. They have maintained that agents' abilities are constituted by a set of dispositions, and that the presence of the counterfactual intervener only masks, or finks, these dispositions. On this view, agents continue to have the ability to do otherwise, even in the presence of counterfactual interveners who would prevent them from exercising this ability.[9] This is not Fischer's view. He holds that, at least in the relevant—all in—sense of 'ability', the ability to do otherwise *is* lost to agents in FSCs. Nevertheless, he maintains that the capacities that underlie the attribution of moral responsibility to agents—capacities to respond to reasons—are not lost. FSCs, like finks and masks, work to focus us on the actual properties of agents and the mechanisms upon which they act, and these—preserved—capacities are what matters to moral responsibility. FSCs are exotic cases, but they enable us to grasp what actually matters to moral responsibility in the normal case

---

[8] For the distinction between finks and masks, see Mark Johnston, "How to Speak of the Colours," *Philosophical Studies*, LXVIII (1992): 221–63.

[9] See Michael Smith, "A Theory of Freedom and Responsibility" and "Rational Capacities," both in Smith, *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics* (New York: Cambridge, 2004), pp. 84–135. For criticism of this approach, see Daniel Cohen and Toby Handfield, "Finking Frankfurt," *Philosophical Studies*, CXXXV (2007): 363–74.

Agents are responsible in FSCs, then, because their actions are the (nondeviant) product of their normal, responsibility-ensuring, capacities Now, obviously, this argument succeeds only if we are entitled to believe that agents in FSCs act on their own in the sense that the presence of the counterfactual intervener does not alter their responsibility-ensuring capacities I shall argue that there is a strong case for thinking that, despite appearances, counterfactual interveners affect the relevant capacities, *even when the intervention remains counterfactual* Rather than arguing (with Smith) that agents in FSCs retain the ability to do otherwise, despite the counterfactual intervention, I shall argue that not only is this ability lost, but so is the capacity to respond to reasons in the normal way, and that therefore the agent cannot be held responsible for her action

As many people have pointed out, Frankfurt's original case can be understood as a clever variation on an example of John Locke's In Locke's thought experiment, a man is carried into a room while he is asleep and the door is locked When he awakes, he finds himself in the company of someone "he longs to see and speak with", as a consequence he stays in the room willingly Locke claims that we cannot doubt that the man stays in the room voluntarily, though "it is evident [ ] he has not freedom to be gone "[10] In an FSC, the agent does not merely lose the ability to succeed in performing an action which he decides not to try performing, he loses the ability even to *try* to perform the action, and yet he seems to decide, on his own, not to try, because he retains the relevant, responsibility-underlying, capacities, in particular the capacities to respond and to react to reasons In what follows, I aim to show that these capacities are lost My strategy will be to follow the pattern laid down by Frankfurt I will begin with an omissions case in which, as in Locke's original case, features of the situation *external* to the agent alter her capacities, then, following Frankfurt's lead, I shall transfer the case inside the head of the agent The cases do not aim at establishing, directly, that agents' capacities can be lost through counterfactual intervention Quite the contrary, I shall show that capacities can be *gained* through counterfactual intervention Demonstrating this point matters because it shows that the strategy that FSCs encourage us to pursue, of understanding agents' capacities by abstracting from counterfactual features of the context, cannot be given a principled rationale If it is true that agents can gain capacities merely by being in the presence of counterfactual interveners, then we are not entitled to conclude that they cannot lose them in the same way, nor are we entitled to claim that understanding agents' capacities requires nothing more than decoding the actual sequence

---

[10] Locke, *An Essay concerning Human Understanding*, II XXI 10

II FRANKFURT ENABLERS

Counterfactual interveners in standard FSCs are counterfactual *disablers*; should the signal for intervention be given, they would intervene to prevent the agent from exercising the responsibility-underlying capacity that might otherwise issue in a choice the intervener aims to prevent In the type of case I shall present, the counterfactual agent is an *enabler* If the conditions for intervention were satisfied, this intervener would act to ensure that the agent is able to perform an action, mental or physical I shall first construct a Locke-style omissions cases in which, due to the presence of a counterfactual intervener, an agent gains a capacity Then I shall proceed to transfer the case into the head of the agent, constructing a Frankfurt-style enabling case (FEC)

I shall begin constructing a Locke-style omissions case by way of a discussion of Fischer and Ravizza's account of responsibility for omissions [11] Their treatment of omissions is symmetrical to their treatment of action cases in both kinds of case, they argue, we must bracket the presence of counterfactual interveners Nevertheless, in omissions cases we cannot ignore the context in which the agent acts, for the following reason responsibility for failing to bring about a particular state of affairs (as opposed, merely, for failing to *try* to bring about such a state of affairs) can only be attributed when that state of affairs is appropriately sensitive to the agent's actions In the absence of such sensitivity, we cannot attribute responsibility for the state of affairs Consider, first, a case of Fischer and Ravizza's in which an agent fails to be responsible for an omission due to some feature of his context

> *Sharks* John is walking along the beach when he notices a child drowning John is a good swimmer, and (psychologically and physically) capable of rescuing the child But John chooses to ignore the child's plight Unbeknownst to John, hungry sharks patrol the water, were he to attempt a rescue, he would be devoured (*ibid*, p 125)

Fischer and Ravizza argue that in this case John is not responsible for failing to rescue the child, because the state of affairs <that the child is rescued> is not appropriately sensitive to John's actions Given the way the world is *actually* constituted, John does not exercise guidance control over that state of affairs, the actual sequence causal route is is saved> But this seems the wrong result I see no reason to think that Jack is not responsible for omitting to save the child He had the right kinds of psychological and physical capacities, and he believed that the state of affairs was appropriately sensitive to his actions Moreover, his belief was *true* had he tried to save the child, he would have succeeded Due to the presence of the Frankfurt-enabler, Jack has gained a capacity he previously lacked the capacity to bring about (or to omit to bring about) the state of affairs <that the child is saved>

---

[11] Fischer and Ravizza, *Responsibility and Control*, pp 123–50

Granted, Jack could only succeed in bring about the state of affairs <that the child is saved> with the help of the good-hearted marine biologist But I see no reason to think that this fact rules out his responsibility for it The case seems to be an unusual case of a common phenomenon joint action When I engage in joint action with another person, with the aiming of bring about state of affairs *S*, and I succeed through this joint action in bringing about *S*, I am at least partially responsible for bringing about *S* What is unusual about Jack's case is that, unlike most ordinary cases of joint action, Jack is unaware that the success of his action depends upon the actions of others But that fact seems irrelevant to the attribution of responsibility to him Consider the case of George, who discovers how to turn off the burglar alarm of a bitter business rival, and decides to turn this fact to his advantage Knowing that if he simply leaves the alarm off all night, the security firm will probably discover the problem and turn it back on, George waits up until he sees a thief about to attempt to break in Seizing his chance, George disables the alarm and the thief makes off with a substantial haul Clearly George is partially responsible for the state of affairs <that his rival is robbed> But, just as clearly, the thief bears at least *some* responsibility for the same state of affairs, even though (unbeknownst to him) he was only able to succeed with George's help Moreover, if the same thief had accepted a financial inducement from George to rob his rival—not knowing, as before, that he could succeed only with George's help—but decides, at the last moment, to return the money and refuse the commission, he might be at least a little praiseworthy for his omission (perhaps there are facts about his circumstances that make it very difficult for him to resist the inducement) Thus, there are action and omission cases in which a state of affairs is sensitive to the actions of an agent only with the cooperation of an enabling intervener, about whom the agent is unaware, but in which we would not hesitate to attribute responsibility for the action or omission to the agent Notice, too, that in the original *Sharks* case, John failed to be responsible due to a factor—the presence of the sharks—concerning which he was unaware Factors of which we are unaware can be relevant to our responsibility for a state of affairs, when these factors would, or would probably, contribute causally to the obtaining or failure to obtain of that state of affairs The ignorance, in *Sharks* and in *CPS* is symmetrical, it would be *ad hoc* to think that it plays a role in one of the cases but not the other

CPS demonstrates, I claim, that we cannot always bracket *counterfactual* features of the environment in which an agent acts when it comes to assessing her responsibility It does this, I further claim, by showing that these—merely counterfactual—features can affect what capacities an agent has But as we have seen, we are entitled to the claim that an agent in a standard FSC is responsible for their action or decision only if we can appropriately bracket counterfactual interveners when assessing agents' capacities If it is illegitimate to bracket counterfactual interveners in cases like *CPS* in order to assess responsibility, then there is a *prima facie* case for thinking that it is illegitimate to bracket counterfactual interveners in FSCs

However, though the agent in *CPS* gains a capacity, it is merely a capacity to bring about a state of affairs An FEC like this leaves the agent *herself* untouched—or so it might be claimed Though *CPS* is bad news for the Fischer and Ravizza analysis of omissions cases, because it shows that the sensitivity of external states of affairs to agents' actions can be affected by counterfactual intervention, such cases leave the larger point—that *agents'* capacities are not affected by such interventions—unchallenged In an FEC, just as much as an FSC, the agent acts *on her own* FSCs force us to focus on the *agent*—specifically, on the reasons-responsiveness of her action-issuing mechanisms—in abstraction from the context (just as finks and masks force us to focus on the structure of entities in order to discover what dispositions they possess) Since an FEC like *CPS* alters only the external sensitivity of states of affairs to the agent's actions, they do not challenge the claim that responsibility depends upon the reasons-sensitivity of the mechanisms that issue in action, and that we test for this sensitivity by bracketing counterfactual interveners

However, just as it proved possible to transfer Locke's original case into the head of the agent, thereby (apparently) showing how internal abilities can be as effectively lost through counterfactual intervention as external, it is possible to construct a "Frankfurt analogue" of *CPS*, that is, an FEC in which the enabling condition external to the agent in *CPS* is transferred into the brain of the agent Such an FEC will have the following features the mechanism upon which the agent acts will be constituted so that it is not moderately reasons-responsive in the actual sequence, because it fails either to be appropriately receptive or to be appropriately reactive to reasons (or both), *but* which has the further feature that *were* the agent to satisfy some condition, analogous to the signal for intervention in an FSC, it would *become* moderately reasons-responsive This might be accomplished either through the action of an intervener, or through some natural process As a consequence, the agent would have *gained* a capacity the mechanism or set of mechanisms upon which she acts would have become either appropriately receptive or appropriately reactive to reasons

Consider this variation on *Sharks*

> *Phobia* Jillian is walking along the beach when she notices a child drowning Jillian is a good swimmer, but she is pathologically afraid of deep water She is so constituted that her phobia would prevent her from rescuing the child were she to attempt to, she would be overcome by feelings of panic Nevertheless, she is capable of *trying* to rescue the child, and she knows that she is capable of trying Indeed, though she knows that she has the phobia, she does not know just how powerful it is, she thinks (wrongly) that she could affect the rescue Unbeknownst to Jillian, a good-hearted neurosurgeon has implanted her with a chip with which he monitors Jillian's neural states, and through which he can intervene if he desires to Should Jillian decide (on her own) to rescue the child, the neurosurgeon will intervene to dampen her fear, she will not panic and will succeed, despite her anxiety, in swimming out to the child and rescuing her

Jillian is psychologically constituted so that the actual mechanism upon which she acts is not moderately reasons-responsive, because it fails to be appropriately *reactive* there is no possible world in which Jillian could bring herself to rescue the child, because in all possible worlds she would be overcome by panic once the water reached a certain depth Now, suppose that Jillian decides not to rescue the child Is she responsible for that omission? This case seems to be closely analogous to *CPS*, and, as in that case, Fischer's injunction to hold the nonintervention fixed seems to yield the wrong answer Jillian thinks she can save the child, and she seems to be right, *despite* the fact that the mechanism upon which she acts in the actual sequence is not moderately reasons-responsive The presence of the counterfactual intervener has given her a capacity that she would otherwise have lacked

If you doubt that Jillian has gained a capacity as a result of the presence of the counterfactual intervener, consider this variation Jillian is psychologically constituted as before she suffers from a paralyzing phobia regarding deep water But the good-hearted neuroscientist has long taken an interest in her welfare, and regularly intervenes to allow her to pursue her interest in scuba-diving Despite the anxiety she feels whenever she enters the water, Jillian has been able to explore many beautiful coral reefs and is a highly proficient diver Surely

Jillian, as now described, is entirely capable of rescuing the child, and deserves condemnation were she to fail to.[13]

Though her actual sequence mechanism is not moderately reasons-responsive, its lack of reasons-responsiveness is irrelevant to what Jillian does, or fails to do. It seems, then, that we can safely set aside the incapacities of her actual sequence mechanism; these incapacities are not *hers*. Here is one way we might put the point: Jillian, or the mechanism or set of mechanisms upon which she acts, consists not only of the phobic machinery, which (left to its own devices) would trigger overwhelming panic in her, but also of the overriding element that ensure that the panic is short-circuited.

Agents in FSCs are, supposedly, responsible for what they do because they do it on their own, in the sense that their decision or action is the product of mechanisms which belong to or which constitute them. They express their, or their mechanisms', unhindered capacities in decisions or actions. But we are entitled to conclude that these agents have their capacities intact only if we are entitled to bracket the counterfactual intervener. As cases like *CPS* and *Phobia* show, we are not so entitled. In FECs agents *gain* capacities, due to the presence of a counterfactual intervener. We can only properly grasp the manner in which the relevant capacities are gained by, precisely, *not* bracketing the counterfactual intervener. The rationale for the claim that agents in FSCs do not lose the relevant, responsibility-ensuring, capacities is therefore undermined. If we do not bracket the intervener in FSCs, in testing agents' capacities, then we are led to the conclusion that these capacities are *lost*. We are, therefore, not entitled to the intuition that agents in FSCs are morally responsible.[14]

---

[13] One possible difference between the two scenarios, which might be felt to be relevant to assessments of responsibility, is that in the latter case Jillian *knows* that she is capable of rescuing the child, whereas in the former she merely has a justified true belief that she can rescue the child. But, even if that is true, I do not see how it is relevant to her responsibility for her omission. Recall the case of George and the burglar. The latter, too, merely had a justified true belief that he was able to rob George's rival; his belief too was only accidentally true and was therefore Gettiered. But the burglar is clearly at least partially responsible for robbing George's rival.

[14] Of course capacities and responsibility can dissociate: an agent can be responsible for failing to ϕ even though she lacks the capacity to ϕ, on the grounds that she can acquire the capacity to ϕ. So we cannot always simply judge an agent's responsibility, or lack thereof, simply by asking whether she possesses the relevant capacities. However, this is no objection to my treatment of FSCs, since it is the opposition—that is, those who believe that FSCs show that agents are morally responsible in FSCs despite lacking alternative possibilities—who are committed to holding that agents are responsible in these cases *because* they retain the responsibility-ensuring capacities. Spelling out just when possession of capacities is responsibility-ensuring is a difficult task; fortunately, it is not a task that I need to engage in here.

III FSCS AND INTERNALISM

There are two ways to resist the conclusion argued for here, that there is no principled reason for bracketing counterfactual interveners when assessing agents' capacities First, we might hope to identify a categorical difference between FECs and FSCs, such that we are justified in treating them in different ways (bracketing counterfactual interveners in FSCs, but admitting their relevance in FECs) Second (and in contrast to this first approach), we might hope to show that there is good reason for bracketing counterfactual intervention in *both* kinds of cases I confess that I am unable to think of an argument for treating FECs and FSCs differently that does not straightforwardly beg the question (by holding that the relevant difference between the cases consists in the fact that in one a capacity is—supposedly—lost, while in the other a capacity is gained) But perhaps such an argument can be constructed Until it is forthcoming, I shall explore the second strategy, for there is a *prima facie* plausible argument that would seem to justify bracketing counterfactual interveners in both FECs and FSCs an argument from internalism about the mental

The lesson of FSCs (as standardly interpreted) is, as we have seen, that it is agents' actual capacities that are responsibility-ensuring FSCs lead us to focus tightly *on the agent* and her mental mechanisms Accepting that cases like scuba-diving Jillian could be relevant to our assessment of responsibility in FSCs requires that we accept that an agent's capacities could be constituted by neural circuitry *plus* some extra-neural, indeed *external*, features But, the objection goes, we have good reason to be (locational) internalists about the mental [15] Just as dispositions like fragility are constituted by the atomic structure of vases or sheets of glass—and not by some world-involving structure, such as the ensemble vase-plus-marble-floor—so agents' capacities are constituted by their neurophysiology Minds supervene on brains, and brains are confined to skulls Now, this kind of internalism about the mental gives us reason to reject the claim, intuitive though it might be, that *Jillian* can scuba-dive Jillian's capacities, the thought goes, are constituted by her neurophysiology But her scuba-diving capacity, we are asked to accept, supervenes on the ensemble Jillian-plus-benevolent-neuroscientist. That *ensemble* might be able to scuba-dive, or perhaps bring it about that Jillian scuba-dives, but *she* lacks the ability to scuba-dive. Whatever our intuitions, in response to cases like scuba-diving Jillian, then, we should set them aside as misleading. In fact, just as proponents of FSCs claim, capacities are constituted by features *of the agent*, and we ought *always* to bracket counterfactual interveners in testing for those capacities.

---

[15] *Locational* internalism is to be contrasted with *content* internalism Content internalists deny that content externalism, of the kind associated with Tyler Burge and Hilary Putnam, is true, that is, they deny that mental states have contents which are partially constituted by features of the environment such that physically identical subjects could have beliefs with different contents Locational internalist holds that mental states are partially *located* external to the skull of the subjects whose states they are, however they get their content

There are two ways to respond to this objection. We might respond by constructing an FEC in which the intervention is internal to the agent. Just as many FSCs involve "blockages" or other elements internal to the skull, so FECs can be constructed in which the counterfactual intervention, or perhaps better, the conditional intervention, is part of the neurophysiology of the agent.[16] The problem with this move is that there is a ready response to it available to—at least some—proponents of FSCs. They can, consistent with their internalism, hold that only FSCs in which the counterfactual intervention is external to the agent are relevant. That is, they can admit that the argument from FECs works to undermine blockage FSCs, by showing that we are not entitled to bracket interventions internal to the agent. But, they can hold, it is always appropriate to bracket interventions—enabling or disabling—external to the agent. Hence, there are, or at least may be, some coherent FSCs in which agents do not lose their responsibility-ensuring capacities; the lessons of FSCs would thus be preserved.

The second, more radical, response to locational internalism about the mental involves confronting it head on. That is, we can *deny* the internalist claim that the mind is confined to the skull. The view, known (variously) as the extended mind hypothesis, environmentalism or locational externalism, according to which mind consists in the set of mechanisms that enable us to perform cognitive tasks, and this set of mechanisms includes external elements—from symbols we manipulate, to marks on paper, to bit and bytes, and even other people—has grown in popularity recently.[17] On this view, mind *does* extend into the world, and, at least on one way of cashing out the extended mind hypothesis, agents, or their actions, can themselves form part of the cognitive mechanisms of other agents

---

[16] For an FSC in which the counterfactual intervener is internal to the agent, see Alfred R. Mele and David Robb, "Blips, Magnets and Seesaws: The Metaphysics of Frankfurt-style Cases," in David Widerker and Michael McKenna, eds., *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities* (Aldershot, UK: Ashgate, 2003), pp. 127–38.

[17] For expositions and defenses, see Andy Clark and David Chalmers, "The Extended Mind," *Analysis*, LVIII (1998): 7–19; Clark, *Natural Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence* (New York: Oxford, 2003); Mark Rowlands, *The Body in Mind: Understanding Cognitive Processes* (New York: Cambridge, 1999); and Robert A. Wilson, *Boundaries of the Mind: The Individual in the Fragile Sciences* (New York: Cambridge, 2004).

The extended mind hypothesis is very controversial However, the radical response to the internalist claim does not depend for its success upon us being able to refute internalism about the mental We need only to show that radical externalism deserves to be taken seriously Recall the dialectical situation On both sides, we have strong and well-entrenched intuitions the principle of alternative possibilities, and its entailments (including, on one plausible reading, the intuitive principle that ought implies can[18]), on the one hand, and internalism about the mental on the other Moreover, the cases, FSCs on the one hand, and FECs on the other, are themselves intuitive (that is, FSCs pump intuitions in just the way their defenders claim, but FECs pump intuitions in just the way I have outlined) Since there are strong and well-entrenched intuitions on both sides, the proponents of FSCs are not entitled to expect the appeal to the intuitiveness of internalism to carry much weight here If defenders of FSCs are to appeal to internalism, they will need to buttress their case with *arguments* for internalism There is a fairly substantial literature defending internalism to which they can appeal [19] Nevertheless (I claim) this is not a promising strategy the internalist arguments are weak, and internalism false

Obviously, I cannot address the arguments for and against internalism here I suspect, however, that a full assessment of the case for FSCs will require properly confronting the case for internalism, for it may well be that the apparent coherence of FSCs stems from internalist intuitions Now, suppose the brain is so constituted that it is naturally a Frankfurt-enabler analogue that is, that agents routinely make decisions using a mechanism that is not appropriately reasons-responsive all by itself, but which, under certain conditions, concatenates with further mechanisms in such a manner as to render the entire ensemble moderately reasons-responsive On this (not unrealistic) picture, agents regularly act on mechanisms that, left to themselves, would always produce a certain output (say, giving in to a desire of a certain strength) but which frequently act as part of an ensemble capable of greater flexibility of output. Now, were the agent to have the link between mechanisms severed, temporarily or permanently, she might retain a capacity indistinguishable, phenomenologically and behaviorally, from her normal capacity to act on her desires, though she would not be able to exercise self-control (because she would no longer be appropriately reasons-responsive). But there would be little temptation to say that she was acting "in the normal way," or that she was responsible for her action, *even if the conditions under which the concatenation of mechanisms would normally occur failed to be satisfied in the particular case.* The belief that she acts in the normal way in an FSC, upon which this case is modeled, is therefore merely an expression of the internalist prejudice: it is only because the counterfactual intervention is external to the agent that we think that we can appropriately ignore it in assessing agents' capacities, specifically their capacity to respond to reasons.

---

[18] For defense of the view that 'ought implies can' must be false if FSCs establish that agents can be responsible in the absence of alternative possibilities, see David Copp, "'Ought' Implies 'Can', Blameworthiness, and the Principle of Alternate Possibilities," in Widerker and McKenna, eds , pp 265–99, and Fischer, "'Ought-Implies-Can', Causal Determinism, and Moral Responsibility," in Fischer, *My Way*, pp 217–22

[19] The best-known defenses of internalism are those of Fred Adams and Ken Aizawa and of Robert Rupert See Adams and Aizawa, "The Bounds of Cognition," *Philosophical Psychology*, XIV (2001) 43–64, and Rupert, "Challenges to the Hypothesis of Extended Cognition," this JOURNAL, CI, 8 (August 2004) 389–428

Human beings routinely gain capacities, including mental capacities, by learning to manipulate external symbols, or by interfacing with features of their environment, including other agents. FECs seem exotic, but they are merely an unusual variation on a ubiquitous phenomenon. But if we gain capacities due to the presence of interveners, even counterfactual interveners, then why can we not lose capacities in the same way? If agents like Jillian can gain in receptivity or reactivity to reasons, then why cannot agents like Connie, in the original voting case, lose such receptivity or reactivity? If they can, then FSCs fail to show that agents can be responsible even in the absence of alternative possibilities. At the very least, I think, proponents of FSCs are not entitled to appeal to the intuition that agents in these cases act in the normal way. They owe us an extra argument to that effect. For the time being, FSCs should not be seen as having broken the dialectical stalemate that characterized the free will debate before they burst onto the scene.

NEIL LEVY

University of Melbourne and
University of Oxford

ON HAVING A POINT OF VIEW: BELIEF, ACTION, AND EGOCENTRIC STATES*

I have all sorts of beliefs about how things are in the world. Many of these beliefs are about how things are in relation to me. I believe not only that the 43rd President of the United States is a Republican or that Philadelphia is east of San Francisco, but also that this piece of paper is white, that there is a table in front of me, and so on. That is to say, my beliefs reflect a particular perspective or point of view. This is a result of the fact that I arrive at my beliefs about the world through perception. I believe that there is a piece of paper on the table because I see it on the table, and when I see the paper on the table, I see it from a certain perspective. Having a point of view is obviously a fundamental feature of human thought and experience. We might nevertheless wonder if this is just a contingent feature of the way we human beings happen to be related to the world our beliefs are about. On the face of it, it seems that we can imagine beings who—unlike us—have a kind of perspectiveless world view: beings who have beliefs about how things are in the world, but no beliefs about how things are in relation to them. My aim in this paper is to suggest that this is more difficult to imagine than one might think, that we can no longer ascribe any thoughts or beliefs about the world to a creature that lacks a point of view.[1]

---

* I have presented earlier versions of this material at University of California, Berkeley, New York University, University of California, Los Angeles, Bryn Mawr College, the 2002 N.E.H. Summer Institute on Consciousness and Intentionality; Wellesley College, University of Edinburgh, Boston University; University of Notre Dame, University of Illinois, Urbana-Champaign, Northwestern University, Johns Hopkins University, Temple University, and Brandeis University. I have benefited greatly from the feedback I received on these occasions. I would also like to thank Bill Brewer, Andrew Brook, Jason Bridges, Tyler Burge, David Chalmers, Mark Greenberg, Gilbert Harman, Sean Kelly, Susanna Siegel, Cara Spencer, Barry Stroud, and Daniel Warren for their helpful comments and advice.

[1] Gareth Evans makes a related claim in *Varieties of Reference* (New York: Oxford, 1982). As Evans puts it, "The idea that there is an objective world and the idea that the subject is *somewhere* cannot be separated" (p. 222). Evans is concerned with what is involved in thinking of one's *perceptions* as perceptions of an objective world: to think of one's perceptions in that way, according to Evans, requires that one "think of his perception of the world as being simultaneously due to his position in the world, and to the condition of the world at that position" (p. 222). See also Evans, "Things without the Mind," in *Collected Papers* (New York: Oxford, 1985), pp. 249–90 (originally published in Zak van Straaten, ed., *Philosophical Subjects: Essays Presented to P.F. Strawson* (New York: Oxford, 1980), pp. 76–119). I am interested in whether we can coherently imagine a subject who has no perceptual experiences—or any other mental state that represents the world from a point of view—at all.

0022-362X/08/0505/240–58

I. EGOCENTRIC STATES

Let me begin by saying something more specific about what having a point of view involves. As I understand it, a point of view involves having what I will call *egocentric states*. Egocentric states are psychological states that represent how things are from the subject's spatiotemporal perspective. Their contents are expressed by first person indexicals or by demonstratives that identify objects or locations in terms of their spatial relation to the subject. The paradigm case of an egocentric state is a perceptual experience. When I see a book on the table, I see the book and table to be in a particular location and orientation with respect to my body. Naturally, perceptual experiences often give rise to egocentric beliefs: when I see the book on the table, not only do I believe that there is a book on a table in such and such a room, I also believe that there is a book on the table in front of me. Other egocentric states include beliefs about the location, position, or posture of the subject (for example, my belief that I am standing up with my hands at my sides), as well as beliefs about which person in the world the subject is (for example, my belief that I am the person standing at such and such a location).

According to this notion of an egocentric state, a belief that it is now raining in Massachusetts does not count as egocentric, despite the fact that it is expressed by the indexical, 'now'. Temporal indexicals alone do not suffice to make a thought or belief count as egocentric. Egocentric states, as I understand them, represent objects from the perspective of the subject's *spatio*temporal location. They reflect the point of view of someone who is embedded in the world as one spatiotemporal object among others. A disembodied mind that experiences the passage of time, but has no perceptual experiences or beliefs about where things are in relation to it might, in some sense, have a point of view, but it is much less robust than the notion that I wish to work with in this paper.

Not all egocentric states require a notion of a self. I need to have a notion of a self in order to believe, say, that I am the person standing at such and such a location. But I do not necessarily require a notion of a self in order to have a perceptual experience of an apple on a table in front of me, or to believe that this apple is red. Some philosophers might insist that all belief and experience presupposes a notion of the self, but even if that were true it would be a feature of thought and experience in general, not something special about egocentric thought and experience. Self-consciousness, then, is not a necessary condition for having egocentric states unless it is a necessary condition for having any psychological state. If it is not a necessary condition, then animals that lack self-consciousness may still have egocentric states to the extent that they have perceptual and other informational states about the locations of objects with respect to their bodies

Self-consciousness is also not a sufficient condition for the possession of egocentric states Egocentric states are, in some sense, more robust than the kinds of thoughts that philosophers sometimes associate with a Cartesian ego While my belief that I am thinking about lunch does have an indexical element, it does not count as egocentric in the way that I am using the term Egocentric states represent the spatiotemporal world from the perspective of a subject in that world My belief that I am thinking about lunch involves a notion of myself as a subject of thought, but it does not involve a spatiotemporal perspective

One final point egocentric states need not be veridical A hallucinatory experience of a table in front of me still represents the table egocentrically, even if in fact no table is there [2] Along these lines, a brain in a vat could have egocentric thoughts and experiences that represent a world from the perspective of a body that appears to be hers

II GLOBAL CLAIRVOYANCE

I will argue that we cannot ascribe thoughts and beliefs about the world to a subject who has no egocentric states This thesis is similar in some respects to Tyler Burge's claim in "Belief *De Re*," that "attributing an understanding of sentences, or propositional attitudes at all, requires attributing *de re* attitudes "[3] For Burge, when we attribute a *de re* belief to someone, we place that person in "an appropriate nonconceptual, contextual relation to objects the belief is about"

---

[2] I am assuming here that hallucinatory experiences and veridical experiences can have similar representational contents This is an assumption that disjunctivists might resist According to disjunctivism, subjectively indistinguishable veridical and hallucinatory experiences are nevertheless different kinds of mental states It is not clear to me whether this means that hallucinations cannot represent objects egocentrically For a defense of disjunctivism, see John McDowell, "Criteria, Defeasibility and Knowledge," in *Meaning, Knowledge, and Reality* (Cambridge Harvard, 1998), pp 369–94 (originally published in *Proceedings of the British Academy*, LXVIII (November 1982) 455–79), or M G F Martin, "The Limits of Self-Awareness," *Philosophical Studies*, CXX (July 2004) 37–89

[3] Tyler Burge, "Belief *De Re*," this JOURNAL, LXXIV, 6 (June 1977) 338–62, see p 348 Another influential version of the thesis can be found in Peter Strawson, *Individuals* (London Methuen, 1959) Strawson argues that we cannot identify spatiotemporal particulars using only descriptions, since it is always possible that a description will be satisfied by more than one object (p 20) Descriptive identification, according to Strawson, must therefore rest on demonstrative identification Bill Brewer defends this claim in *Perception and Reason* (New York Oxford, 1999), pp 18–48 I raise some worries about the argument in "The Threat of Massive Reduplication" (manuscript)

(*op cit*, p 346), it is "nonconceptual" in the sense that "the believer's relation to the relevant object is not merely that he conceives of it or otherwise represents it" (*op cit*, pp 346–47) Burge says that a sufficient condition for an attitude to be *de re* "is for it to contain an indexical expression used deictically" (*op cit*, p 347) While it may be that not all *de re* attitudes are egocentric, egocentric states are paradigmatic cases of *de re* attitudes When I believe that this book is blue, I am standing in exactly the sort of "contextual relation" Burge has in mind

Burge wants to argue against the possibility that "some organism or robot could be programmed to understand an indexical-free language, without our attributing any *de re* attitudes to it" (*op cit*, p 347) What distinguishes someone who understands a language from a mere symbol manipulator, according to Burge, is an ability "to recognize and initiate correlations between symbols and what they symbolize" (*op cit*, p 347) For example, the speaker might pick up a ball when someone tells her to get a ball or say "There is a ball" when a ball is in front of her (*op cit*, p 348) Burge argues that if a speaker exhibits that sort of behavior, we will naturally ascribe *de re* attitudes to her We will conclude, for example, that she believes that someone wants her to fetch *that ball* So if we were not in a position to ascribe *de re* attitudes to a subject, we would not be in a position to ascribe any propositional attitudes to the subject at all

Burge's "organism or robot" is similar to the Turing machine that Hilary Putnam considers in the first chapter of *Reason, Truth, and History* This machine lacks "electronic eyes and ears" and has "no provisions in the machine's program [ ] for incorporating inputs from such sense organs, or for controlling a body "[4] Putnam says

> To me, it seems evident that we cannot and should not attribute reference to such a device It is true that the machine can discourse beautifully about, say, the scenery in New England But it could not recognize an apple tree or an apple, a mountain or a cow, a field or a steeple, if it were in front of one (*ibid*)

> Our talk of apples and fields is intimately connected with our *non-verbal* transactions with apples and fields There are 'language entry rules' which take us from experiences of apples to such utterances as 'I see an apple', and 'language exit rules' which take us from decisions expressed in linguistic form ('I am going to buy some apples') to actions other than speaking Lacking either language entry or language exit rules, there is no reason to regard the conversation of the machine [ ] as more than syntactic play (*ibid*, p 11)

---

[4] Putnam, *Reason, Truth and History* (New York Cambridge, 1981), p 10

I do not think these considerations are adequate on their own to explain why egocentric states are indispensable for the possibility of thought. Both Burge and Putnam ask us to consider an organism or machine for which there are no connections between the symbols it manipulates and objects or events in the world. But one might think it conceivable for symbols to be connected with objects that are not identified egocentrically. To see what I mean, try to imagine the following case. Imagine a subject who has no egocentric states at all. She has no perceptual experiences and no thoughts, beliefs, or desires that contain any demonstrative or indexical elements. She is nevertheless caused to have all sorts of true (nonegocentric) beliefs about the world through a nonperceptual mechanism. This subject forms all of her beliefs in the way that a clairvoyant might come to believe that a particular event—say, a murder or a natural disaster—has occurred.[5] Admittedly, clairvoyants are often said to have visions of the events in question. A clairvoyant might "see," in some sense of the word, an image of the place where a murder victim is buried. But we can also easily imagine—and perhaps have even heard of—clairvoyants who just find themselves with a strong (and accurate) conviction that something terrible has happened; the causal relation between the conviction and the event in question is not mediated by visions of any kind. These people are not plagued by images of the misfortune; they just find themselves believing that the tragedy has occurred. This is how we should imagine our subject. She has no visions or perceptual experiences; instead she just finds herself with all sorts of beliefs about the world. I will call this a case of "global clairvoyance." The purpose of the term "global" is to distinguish this kind of clairvoyance from the cases we tend to hear about: cases in which people form some beliefs through nonperceptual channels, but arrive at the bulk of their ordinary beliefs through perception.

We can suppose that our global clairvoyant speaks: when an event causes her to adopt a belief, she utters a sentence that corresponds to the event in question. So whenever it is raining in Massachusetts, for example, she utters the sentence "It is raining in Massachusetts"; whenever another species of butterfly goes extinct, she utters, "Another species of butterfly is now extinct," and so on. Since her utterances are reliably correlated with what is going on in the world, the global clairvoyant would be quite useful to those who cross paths with her.

---

[5] I am indebted here to Sydney Shoemaker's discussion of clairvoyance in *Self-Knowledge and Self-Identity* (Ithaca: Cornell, 1963), pp. 174–82. Shoemaker thinks that the ability to make "egocentric statements" is what distinguishes genuine perception from clairvoyance (pp. 175–76). The case also bears some resemblance to David Lewis's example of the two gods in "Attitudes *De Dicto* and *De Se*," *The Philosophical Review*, LXXXVIII (October 1979): 513–43.

Clearly, there are *connections* between the global clairvoyant's utterances, beliefs, and the objects and events they are about, we are supposing, after all, that the beliefs are formed through a reliable causal mechanism. And yet we are supposing that none of these beliefs is egocentric. The clairvoyant lacks a point of view. While she has many true beliefs about how things are in the world, she has no beliefs about how things are in relation to her. Nothing in Putnam's content externalism rules out the possibility of global clairvoyance. The crucial point for externalists is that a person's beliefs are about the objects and events that regularly cause them. As it happens, we are caused to form beliefs through perception, as such we have a rich body of egocentric states that constitute a point of view. But why think that the way in which we are typically caused to form beliefs is the only possible way? Content externalism on its own seems to have no answer to this question.

Is there anything in Burge's argument that would rule out the possibility of global clairvoyance? I can think of two possible responses one might make on Burge's behalf. First, one might respond that while the global clairvoyant's beliefs are not egocentric, many of them are still in fact *de re*. This would allow for the possibility that the global clairvoyant could have beliefs, without threatening Burge's claim that *de re* beliefs are essential for propositional thought. Our ability to ascribe *de re* beliefs to the clairvoyant will depend on whether there is "an appropriate nonconceptual, contextual relation" between her beliefs and the objects or events they are about. One might argue that the relation is "nonconceptual" in the sense that it consists in more than the mere fact that the clairvoyant "conceives of" or "represents" the objects her beliefs are about. she has a causal, albeit nonperceptual, connection to those objects. In any case, whether the clairvoyant has *de re* attitudes or not, the beliefs we are interested in are *egocentric* attitudes that reflect a particular perspective or point of view. Even if some of the clairvoyant's beliefs count as *de re*, we are still left with the question of whether *egocentric* states are required in order to have any thoughts about the world at all.

The second response attempts to show that the global clairvoyant, as we have imagined her, cannot possibly have any propositional attitudes. Perhaps Burge could insist that mere causal connections are not enough for propositional thought. In addition, a person must "recognize" the connections between her propositional attitudes and the things they are about. There is something intuitive about this line of thought. However, much more still needs to be said: we need to articulate more clearly why this sort of recognition is required or, alternatively, why mere causal connections do not count as cases of recognition. Why is a reliable capacity to form true beliefs not enough to count as an ability to recognize connections between those beliefs and the objects they are about? And if it is not enough, then why is recognition—as opposed to the reliable formation of true beliefs—a requirement for having propositional attitudes in the first place?[6]

I do think there is something suspect about the notion of global clairvoyance. Burge is on to something important when he points out that the abilities a person must have in order for us to ascribe any propositional attitudes to her at all are the very same abilities that would also lead us to ascribe *de re* attitudes. Nevertheless, we still need to explain why these abilities involve having the kinds of egocentric states that constitute a point of view. In what follows, I will argue that if we suppose that none of the clairvoyant's mental states are egocentric, her perspectiveless world-view becomes unrecognizable as a body of beliefs. The argument begins in section III with the claim that we must be able to ascribe egocentric states to a person in order to understand her bodily movements as intentional actions. I will then make a similar point in section IV about the clairvoyant's utterances: to the extent that the clairvoyant lacks egocentric states, her utterances cannot be understood as intentional. As a result, we cannot regard those utterances as expressions of thoughts and beliefs; her utterances will be more akin to the level of mercury on a thermometer than to meaningful speech. In section V, we will see that on a reasonable understanding of what it is to be a mental state of a given kind, the clairvoyant's mental states have so little in common with genuine thoughts and beliefs that it becomes difficult to think of them as such.

---

[6] These issues are closely related to a debate between McDowell and Robert Brandom about whether the reports made by "chicken-sexers" count as observational judgments. The chicken-sexers are supposed to have a remarkable ability to tell whether a baby chick is male or female even though to them (and to everyone else) the chicks look exactly alike. Brandom thinks that the chicken-sexers' ability to differentiate reliably between male and female chicks is enough for their reports to count as observations (provided the responses can serve as grounds for further inferences and commitments). McDowell, by contrast, argues that the chicken-sexers' reports cannot count as observations since the chicken-sexers are not responding to how things look. See Brandom, "Non-inferential Knowledge, Perceptual Experience, and Secondary Qualities: Placing McDowell's Empiricism," in Nicholas Smith, ed., *Reading McDowell: On Mind and World* (New York: Routledge, 2002), pp. 92–105, see pp. 97–98, and McDowell's response (pp. 279–81).

III EGOCENTRIC STATES AND INTENTIONAL ACTION

In "The Problem of the Essential Indexical,"[7] John Perry makes a persuasive case for the idea that we must use indexicals to characterize certain beliefs that are required to give a rational explanation of a subject's bodily actions "These indexicals are essential," claims Perry, "in that replacement of them by other terms destroys the force of the explanation, or at least requires certain assumptions to be made to preserve it" (*ibid*, p 28) To use Perry's well-known example, a proper explanation of why I stop and reach for the torn bag of sugar in my grocery cart must appeal to my realization that the person spilling sugar all over the aisles of the grocery store is *me* (*ibid*, p 27) The explanation ceases to have any force if we replace 'me' with some nonindexical co-referring expression While this point is widely recognized, I believe that less attention has been paid to its full consequences for our conception of intentional action and thought

Perry is merely trying to argue that there are *some* cases in which understanding why a person acts the way she does requires citing beliefs that can only be expressed using indexical terms But I think that we can extend Perry's point to bodily actions generally Any rational explanation of a bodily action must involve the ascription of egocentric beliefs How, for example, can we understand why a subject reaches over to her left to pick up a coffee mug unless we take her to believe that the mug is on her left? Naturally we must also take her to have particular desires or intentions presumably she would not have reached for her mug unless she wanted to drink from it (or put it in the sink, or refill it, and so on) But the ascription of desires and intentions alone does not suffice to explain bodily action Agents generally perform bodily actions so that they can effect changes in the environment in order to satisfy some desire A proper explanation of such actions should therefore specify not only the desire the agent is trying to satisfy, but also where the agent takes the relevant objects to be located in relation to her[8] Otherwise, it is inexplicable why the agent moves in one direction rather than another Clearly, then, the subject's egocentric states play an essential role in making sense of her behavior Without them, her behavior is inexplicable to us

---

[7] Perry, "The Problem of the Essential Indexical," in *The Problem of the Essential Indexical and Other Essays* (Stanford CSLI, 2000), pp 27–44 (originally published in *Noûs*, xiii (March 1979) 3–21)

[8] This claim may require some qualification One might object that the agent does not need to have beliefs about the locations of objects in relation to *her*; concepts such as *this body* or *here* could also do the same work That is to say while egocentric states are required to explain action, such states might not need to employ a first-person concept This raises large issues about first-person thought that I do not have room to address here (A fuller account might look at how action explanations would work if a subject received sensory information from—and could control the movements of—

In "Demonstrative Thought and Psychological Explanation,"[9] Christopher Peacocke appeals to similar examples to motivate what he calls the "Indispensability Thesis " He states the thesis as follows

> No set of attitudes gives a satisfactory psychological explanation of a person's acting on a given object unless the content of those attitudes includes a demonstrative mode of presentation of that object (*ibid*, pp 205–06) [10]

While the Indispensability Thesis is close to what I want to argue for, it differs in two respects First, Peacocke's claim is restricted to a person's actions on a particular object However, not all actions are actions on an object walking to the store, raising one's hand, and dancing are all intentional actions, but it is not obvious what particular object they are actions on Second, Peacocke requires that the person's attitudes include a *demonstrative* mode of presentation of the object on which the agent is acting An explanation of why I reach out to grasp a coffee mug in front of me must appeal, for example, to a visual experience I have of *that mug* in front of me or a belief that *that mug* is filled with hot coffee

My thesis is both weaker and broader than Peacocke's thesis I am claiming that any adequate psychological explanation of a person's intentional bodily action involves the ascription of egocentric states This claim is weaker in the sense that it does not require a demonstrative mode of presentation of a particular object While many egocentric states do contain demonstrative modes of presentations, a mental state does not have to contain a demonstrative presentation of an object in order to be egocentric For example, while my belief that there is a mug roughly two feet to my left does not contain a demonstrative mode of presentation of the mug, it nevertheless counts as egocentric since it identifies the mug in terms of its re-

---

another body See, for example, Daniel Dennett, "Where Am I?" in *Brainstorms* (Cambridge MIT, 1981), pp 310–23, or Evans, *Varieties of Reference*, p 221) In any case, this does not affect the larger point I want to make about the indispensability of *egocentric* states (not just states that employ a first-person concept)

[9] Peacocke, "Demonstrative Thought and Psychological Explanation," *Synthese*, XLIX, 2 (November 1981) 187–217

[10] The thesis is taken up again in Peacocke, *Sense and Content Experience, Thought, and Their Relations* (New York Oxford, 1983), pp 158–60

lation to me[11] My thesis is broader than Peacocke's in that it concerns any intentional bodily action, not merely actions on particular objects As such it applies to intentional actions whose descriptions do not make reference to any objects in the vicinity of the agent I have in mind actions such as raising one's arm (I am assuming here that the agent is not raising her arm in order to reach something, and so on), tapping one's toes, snapping one's fingers, dancing, and so on[12] In performing these actions, the agent moves her body, but not for the sake of affecting any changes on the things in her environment To explain such actions, we might not need to appeal to her attitudes about the objects around her Nevertheless, we must still assume she has an intention to move her body in a certain way, as well as background beliefs or experiences concerning the current position of her body

It is reasonable to assume that if it is not possible to give a rational explanation of her behavior, we cannot understand anything she does as an intentional action Part of what distinguishes intentional actions from mere bodily movements is that intentional actions are explicable in terms of the agent's beliefs and desires That is to say, intentional actions—unlike mere bodily movements—are performed for reasons If this understanding of intentional action is correct, then if it is not possible, even in principle, to explain why a person acts the way she does, it becomes difficult to see what she does as an action and not just purposeless movement on her part The implications for the global clairvoyant should be obvious once we assume that the clairvoyant has no egocentric beliefs, we no longer have the resources to explain why she behaves the way she does And if it is not possible to give a rational explanation of her behavior, we cannot understand anything she does as an intentional action[13]

IV EGOCENTRIC STATES AND LINGUISTIC UTTERANCES

I have concentrated so far on the clairvoyant's bodily behavior I would now like to turn my attention to the clairvoyant's verbal behavior Not surprisingly, the points I have been making about agency also hold here. Verbal behavior is, after all, a kind of bodily behavior: when a person speaks she moves her lips and whatever else (tongue, vocal chords) is required for the production of sound. But unlike a lot of our bodily movements (with the exception of sign language, hand signals, and so on), the utterances we make often have meanings. For that reason, by applying what I have argued so far to the clairvoyant's verbal behavior, we get some important results: it turns out that we cannot take her utterances to be meaningful expressions of any thoughts or beliefs. At best, they are merely *symptoms* of various states of affairs.

---

[11] Peacocke argues that beliefs containing descriptions such as "roughly two feet" can explain action only if one's understanding of the length of a foot has a demonstrative component for example, "two feet is *this* long" ("Demonstrative Thought and Psychological Explanation," pp 207–08)

[12] Perhaps Peacocke could insist that such actions are actions on objects—namely, one's own limbs—that are presented to the agent demonstratively

[13] In *The Subjective View*, Colin McGinn claims, for similar reasons, that a subject with no egocentric states would be "deprived of the power to act" (New York Oxford, 1983), p 104 McGinn goes on, however, to deny that this explains the "ineliminability" of indexical thought According to McGinn, "the obvious question is whether it is possible for a non-agent to be innocent of egocentric representations" (p 104) My aim in the rest of the paper is to suggest some difficulties with this possibility

We are supposing that the clairvoyant makes utterances that are correlated with various events: when an event causes her to adopt a belief, she utters a sentence that corresponds to the event in question. Would we be able to make sense of her utterances as actions? In other words, could we see her utterances as intentional, as something she intends to perform? I think it would be a stretch to understand her utterances in that way. Like other bodily movements, a person's utterances can be understood as intentional actions only if we assume that she has egocentric states. This point becomes clear if we reflect again on what it takes to understand something as an intentional action. We cannot see something as an intentional action unless it is at least possible to explain why the person performs it. To apply this idea to a person's utterances, we cannot see these utterances as intentional unless it is possible to explain why she makes them or—to put the point another way—what the subject is doing in making those utterances. What we need to see, then, is whether the clairvoyant's lack of egocentric states prevents us from understanding what she is doing when she speaks. There are several things people typically do when they make utterances. They speak to impart information, make promises, make requests or commands, and so on. The examples I have listed so far all involve having some sort of effect on another person. When I speak to impart information my purpose is generally to get someone to believe something. When I make a request, my purpose is often to get someone to do something. This is just to say that, in many cases, communication is the purpose of speech. And this involves having an effect on other people by means of the words we utter.

I want to argue that we cannot understand a speaker as intending to communicate anything unless we suppose that she has at least some egocentric states. It is inexplicable how a speaker could intend to have an effect on a hearer by means of her utterances if she has no understanding of where the hearer is located with respect to the source of these utterances The source of a speaker's utterances is, of course, her body (more specifically, her mouth and vocal chords) This is true, at least, with face-to-face communication There are, of course, less direct methods of communication where the speaker may not know the exact location of her hearer I can talk to someone on the telephone without knowing exactly where she is But even in these cases, egocentric states play an essential role It is not clear how the speaker could intend to have an effect on her audience by means of her utterances if the speaker had no understanding of where the telephone receiver is located with respect to her own body In this way, communicative speech is more akin to other bodily actions than one might think Both kinds of acts involve using one's body to have some effect on the environment It is just that in communication, what we are trying to affect in our environment is another person Moreover, the intended effect is often psychological we want the other person to believe something, to feel a particular emotion, and so on These differences, however, do not change the fact that we attempt to produce such effects by acting on our surroundings We make noises that we hope the other person will hear and understand This involves having some beliefs about where the other person is in relation to us, or perhaps—as in the case of telephone calls—where our body is in relation to the receiver that will carry our voice to the person on the other end of the line

Since the global clairvoyant has no egocentric beliefs, she has no beliefs about where other people (or telecommunication devices) are located in relation to her body This prevents her from understanding how the utterances she makes could affect another person As a result, I think it is difficult to view her utterances as attempts on her part to affect anyone, and it is therefore puzzling how she could be acting intentionally when she makes those utterances I do not mean to insist that a speaker must have very specific beliefs about the physical processes by means of which communication takes place I do not need to understand completely how telephones work in order to intend for my utterances into the receiver to have an effect on the person I am calling Nor must I have any detailed beliefs about the physiology of speech My point is merely that if something a person does is to count as an intentional act, the person must have at least some understanding, however vague, about how her action might accomplish what she intends to do Otherwise there is no rational explanation of why the person does one thing rather than another And this is exactly what is lacking in a person who has no egocentric states If a person has no conception of where anything is in relation to her body, then she has no conception of how what she does with her body—whether it is moving her limbs or her lips—affects anything around her. In that case, it becomes quite a stretch to understand any of her physical or verbal behavior as an intentional action. Intentional actions are distinguished from unintentional bodily movement precisely because actions are performed for reasons. The global clairvoyant is missing the beliefs necessary for her to have any such reasons.

I have argued so far that we cannot take the clairvoyant's utterances as attempts on her part to affect anyone. In other words, we cannot view her utterances as acts of communication. But perhaps one could object that not all intentional utterances are acts of communication. After all, people sometimes talk to themselves when no one else is around, and we would hesitate to conclude that such speech is unintentional. Nevertheless, it is difficult to make sense of the clairvoyant intentionally talking to herself. The clairvoyant cannot intentionally speak, even if it is only to herself, if she lacks the intention to do so. And without any egocentric states, the clairvoyant cannot have any such intention: she cannot intend to emit sounds from her mouth. In this way, talking to oneself is akin to other kinds of actions that we can describe without making reference to any objects (in this case, other people) in the agent's environment. In all such cases, one at least needs an intention to perform the action in question—and such an intention is irreducibly egocentric.

I have been trying to show how difficult it is to find bodily agency in a person who has no egocentric states. This goes not only for bodily movements such as walking across the room or raising one's hand, but also for the verbal utterances one makes. In fact, my purpose has been to bring out the fact that speaking is a bodily activity like any other and, like other bodily actions, requires having some egocentric states. In the absence of such states, it is inexplicable why the clairvoyant makes any of her utterances, and so her utterances cannot be understood as intentional acts on her part. This has an interesting effect on how we should understand the meaning of the clairvoyant's utterances. We can see the effect most clearly if we look at Paul Grice's distinction between natural and non-natural meaning.[14] To use one of Grice's examples: when the doctor says, "Those spots mean measles," she is using 'mean' in the natural sense. Nobody means anything by the spots. It is just that the spots on the patient's skin are a reliable sign that the patient has measles, there is a causal connection between the appearance of spots and the presence of measles On the other hand, "That remark, 'Smith couldn't get on without his trouble and strife', meant that Smith found his wife indispensable," is an example of non-natural meaning What makes it non-natural is the fact that someone meant something by that remark, someone made that remark in order to convey that Smith could not get by without his wife

---

[14] Grice, "Meaning," *Philosophical Review*, LXVI (July 1957): 377–88.

Grice characterizes non-natural meaning as follows In order for a speaker to mean something by an utterance, he must intend to induce a particular belief in his audience by means of that utterance He must also intend for his audience to recognize that he intends to induce that belief by making the utterance, the audience's recognition of his intentions is supposed to play a part in inducing the belief in question For example, when Smith's colleague says, "Smith couldn't get on without his trouble and strife," he intends his audience to come to believe that Smith finds his wife indispensable, and he thinks that his audience will come to adopt this belief by recognizing his intention to induce it (*ibid*, pp 383–84) Philosophers argue about the adequacy of this characterization of non-natural meaning, and it is not my purpose to enter the details of that debate here But I think that the general idea is very plausible what distinguishes natural from non-natural meaning has to do with an intention on the part of the speaker To say that someone meant something by a remark is to imply, at the very least, that the person intended to make that remark A key difference between the man who says, "It is hot outside" and the level of mercury on the thermometer is that the remark is intentional The thermometer does not intend to do anything And we can accept this basic point without committing ourselves to the details of Grice's account

It should now be clear that the clairvoyant's utterances seem more like the level of mercury on the thermometer or the spots on the patient's skin than the man's remark about Smith and his wife On the assumption that there are causal regularities between the clairvoyant's utterances and events in the world, we can use her remarks as reliable indicators of various occurrences in the way that the level of mercury in the thermometer is a reliable indicator of the temperature But we cannot say that she means anything by the remarks she makes We are barred from understanding her utterances in that way precisely because we are barred from thinking of them as intentional actions To mean something by a remark requires, at the very least, that one intends to make it

I am obviously not the first to invoke the capacity to act in an account of what it takes for a speaker's utterances to be expressions of thoughts and beliefs In particular, this discussion might remind some of an old debate between Norman Malcolm[15] and H. H. Price[16] about whether inanimate objects could have psychological states. Price argues that if a tree frequently makes sounds that resemble true sentences, and if those sentences sometimes tell us things about the world we did not already know, we would have reason to believe that it "was animated by an intelligence like [our] own" (*ibid.*, p. 436). (The emphasis on the newness of the information is to rule out the possibility that the utterances came from us.) For Price, we need not see the source of the utterances as an agent in any other respect; in fact, the utterances need not proceed from a body at all. Presumably, Price would think that we have plenty of evidence that the global clairvoyant has thoughts. The clairvoyant makes sounds that resemble intelligible speech, and we could use those utterances to find out things we did not already know.

Malcolm, by contrast, insists that "no amount of intelligible sounds coming from an oak tree or a kitchen table could create any probability that it has sensations and thoughts" (*op. cit.*, p. 973). What is missing, according to Malcolm, is the ability to exhibit the kind of behavior that would show that the tree or table "*understands* the sounds that come from it" (*op. cit.*). A tree might be able to "define" words (for example, 'vixen') in terms of other words (for example, 'a female fox'). What it cannot do is to act in a way that suggests it can "make correct *applications* of the word in question" (*op. cit.*). The behavior Malcolm has in mind involves "looking, pointing, reaching for and going to the right things and not the wrong ones" (*op. cit.*). We determine that a child understands the word 'blue' when she points to blue things and says 'blue', when she reaches for blue things when we ask her to, and so on. Since trees cannot point, fetch or reach for things, Malcolm says that they "cannot either pass or fail the tests that a child is put through. They cannot take them" (*op. cit.*, p. 974). In response, we might wonder why Malcolm thinks that the tree's ability to utter true sentences does not count as an ability to "make correct *applications* of the word in question." Suppose the tree makes accurate statements about the color of the tie that President Bush is wearing that day. Why would that not amount to an ability to make correct applications of words like 'red' and 'blue'? Why must the tree in addition have the ability to point to or reach for red and blue things?

---

[15] Malcolm, "Knowledge of Other Minds," this JOURNAL, LV, 23 (November 6, 1958): 969–78.

[16] Price, "Our Evidence for the Existence of Other Minds," *Philosophy*, XIII, 52 (October, 1938): 425–56.

Malcolm relies on the assumption that the ability to exhibit certain kinds of behavior is a *criterion* for linguistic understanding (*op cit*, pp 974–95) The connection I wish to make between agency and linguistic understanding is different and, I hope, less controversial My starting point is this to be the expression of a person's belief, an utterance must count as an intentional action A person must be an agent in order for the sounds she makes to have non-natural meaning, and this in turn requires having egocentric states Suppose for a moment that we could not interpret any of the sounds that emanate from Price's tree as egocentric statements, and that the tree does nothing else to suggest that it has any beliefs or experiences that represent objects egocentrically The tree has no point of view In that case, there would be no rational explanation of why the tree is making those sounds and we could not see those sounds as intentional actions For that reason, the sounds the tree makes could not be expressions of thoughts and beliefs

I agree with Malcolm that the ability to point at or reach for objects is an important factor when we determine whether the sounds a creature makes indicate the presence of intelligence But for me, such abilities are important, not merely because they would show that the creature can correctly apply the most basic words in its vocabulary, but because they would indicate that the creature has egocentric states If someone reaches out and picks up an apple, we are inclined to say that she believes, or sees, that there is an apple in front of her

V DOES THE CLAIRVOYANT HAVE BELIEFS?

Our global clairvoyant is beginning to look quite impoverished Nothing she does can be understood as an action There is no rational explanation for any of her bodily movements Any utterances she happens to make could have only natural meaning, assuming they have any meaning at all Even if we find causal regularities between her utterances and events in the world, we cannot say that she means anything by her remarks In this sense, her utterances seem more like the readings on a thermometer than expressions of beliefs Can we still hold on to the idea that the clairvoyant has beliefs? In light of all of the things we are barred from ascribing to the clairvoyant, I think it is a difficult idea to maintain Whatever mental states we supposed she had are entirely unconnected to anything else about her Her beliefs are like unconnected cogs in a system of gears, they have no recognizable rational role to play at all All we have here are mental events that are regularly caused by various events in the world It is tempting to say that these occurrences in the clairvoyant's mind are not all that different from the utterances that accompany them Her states have "content" in the sense that they indicate certain states of affairs just as her utterances have "meaning" in the sense that they are symptoms of the events they accompany

Whether we still call these mental occurrences beliefs depends, of course, on what sorts of things we take beliefs to be Here is one widely held view about the nature of belief what it is to be a belief must be understood, at least in part, in terms of the rational role that beliefs play in a person's psychology And the role of belief is intimately connected with action Our beliefs, in conjunction with our desires, give us reasons for action [17] This does not mean that every belief must be a reason for an action a person performs in her lifetime in the case of more theoretical beliefs, the connection with action may be quite indirect It does mean—at the very least—that a subject cannot count as a believer unless she is also a rational agent That is to say, a believer must be the sort of subject who has reasons for acting, even if she does not perform those actions on particular occasions [18] Lacking egocentric beliefs, the clairvoyant has no such reasons, she is not a rational agent And so nothing that goes on in the clairvoyant's head could help explain anything she does If this picture of belief is correct, it will be a stretch to call the clairvoyant's states beliefs

I can imagine two lines of resistance to this picture of the relation between belief and agency The first is to reject that idea that a belief is something that plays a particular kind of role in a person's rational psychology, insisting instead that each mental state has an intrinsic qualitative character This qualitative character makes that state count as a belief, desire, and so on, without reference to any connections the state has to anything else about the subject I do not have space to argue against this view here, though I can say that it is a picture of mental states that few philosophers would feel comfortable embracing [19]

---

[17] See Robert Stalnaker, *Inquiry* (Cambridge MIT, 1984), p 19, for a helpful discussion of this picture of mental states For somewhat different reasons, the view that cognition depends on action is also a central theme in the "embodied cognition" movement in the philosophy of mind and cognitive science See, for example, Andy Clark, *Being There Putting Brain, Body, and World Together Again* (Cambridge MIT, 1997)

[18] Someone who suffers from total paralysis may have reasons for acting—assuming her perceptual faculties are intact—even though she is physically incapable of performing the actions in question She may, for example, have a reason to reach over for the glass of water she sees next to her, even though she cannot in fact pick up the glass The clairvoyant, on the other hand, cannot even have a reason for reaching over for the glass, since she lacks any beliefs about where the glass is located with respect to her body

[19] For an exception, see David Pitt, "The Phenomenology of Cognition or What Is It Like to Think That P?" *Philosophy and Phenomenological Research*, LXIX, 1 (July 2004) 1–36

The second line of resistance is to accept the idea that beliefs must be understood in terms of the role they play in a person's psychology, while denying that this role must involve giving reasons for action. One alternative is to appeal to rational connections between beliefs. The relevant role of belief, in this case, is not to provide reasons for action, but rather to enable inferential transitions to other beliefs. Beliefs would essentially be positions in what Wilfrid Sellars calls the "space of reasons."[20] (The difference is that there need not be any "language-exit" moves from beliefs to bodily actions.) Let us suppose, then, that the clairvoyant could make inferences from one belief to another. Perhaps she uses the information she receives through clairvoyance to make predictions about what will happen in the future, or perhaps she constructs scientific theories on the basis of that information. That would be enough, on this account, for the clairvoyant's mental states to count as genuine beliefs; there is no further need for those beliefs to give her reasons for acting in certain ways.

The problem with this line of thought is that it lacks the resources to distinguish between genuine inferential transitions between beliefs and casual transitions between nonpsychological representational states. We can imagine a more sophisticated thermometer that computes the heat-index by combining data about the temperature and humidity. There is a sense in which it makes "inferences" from one representational state to another. It is roughly the same sense in which the spots on a patient's skin "mean" that the patient has measles. Just as we distinguish between natural and non-natural representational states, we can also distinguish between natural and non-natural "inferences" from one representational state to another. In this way, the problem of distinguishing the clairvoyant's states from the states of a thermometer returns as the problem of distinguishing the clairvoyant's so-called "inferences" from the merely causal transitions between the representational states of the thermometer. If the transitions the clairvoyant makes play no role in giving her reasons for action, are we still entitled to view them as genuine inferences? I think more needs to be said here, but I doubt that mere transitions between states are rich enough on their own to support the idea that the clairvoyant has a genuine mental life. If these transitions are not connected with reasons for action, I suspect that we lose the right to think of them as inferences between beliefs.

---

[20] Sellars, *Empiricism and the Philosophy of Mind* (Cambridge: Harvard, 1997), p. 76. This notion is developed in different ways in McDowell, *Mind and World* (Cambridge: Harvard, 1996), and Brandom, *Making It Explicit* (Cambridge: Harvard, 1994).

We are now in a position to appreciate the importance of having a point of view on the world our beliefs are about. I have tried to argue that if we take seriously the idea that the clairvoyant has no egocentric states, we are hard-pressed to understand her as having any beliefs at all. This is because without such states, we cannot see any agency in the clairvoyant's behavior. If nothing she does can be understood as an intentional action, the clairvoyant's mental states have no role to play in her rational psychology. On one natural understanding of what it is to have a belief, it seems forced to think of the clairvoyant's representational states as beliefs, as opposed to mere symptoms of the states of affairs in the world with which they are correlated. She just seems too impoverished to have much of a mental life. Perhaps the global clairvoyant would be very useful as a tool for us to determine what is going on in the world. We could consult her instead of watching the evening news. But however useful she may be, she would still be more like a thermometer than a genuine person.

If externalism is correct, then we cannot have beliefs about the world unless there are causal connections between at least some of those beliefs and the objects or events they are about. But one might think that it does not matter much what kind of causal connections those are. While we human beings happen to be caused to form beliefs through perceptual channels, one might think that things could have been otherwise. If what I have said about the global clairvoyant is correct, however, then it turns out that not any cause will do. Having any beliefs about the world at all involves having perceptual faculties, or at the very least faculties that closely resemble perception by delivering information perspectivally. Such faculties would have to cause the subject to have the egocentric states required for intentional action. Some might think that this would just be a case of perception, that perception is nothing more than a source of information about how things are in relation to one's body.[21] Others might insist that perception requires having perceptual *experiences*. I happen to think that one cannot have a point of view without having perceptual experiences, but I will not argue for that claim here.

CHERYL K. CHEN

Harvard University

---

[21] See, for example, David Armstrong, *Materialist Theory of the Mind* (New York: Routledge, 1993), pp. 208–39.

COMMENTS AND CRITICISM

DEONTOLOGY, INDIVIDUALISM, AND UNCERTAINTY
A REPLY TO JACKSON AND SMITH*

Discussions of deontological moral theories typically focus on the advantages and disadvantages of deontological constraints, rules to the effect that some actions should not be performed—at least sometimes—even when performing them will maximize the good. And, of course, the jury is still out on whether deontological constraints can be defended.

But in their recent paper "Absolutist Moral Theories and Uncertainty," Frank Jackson and Michael Smith[1] emphasize not the general and well-known challenges to deontological constraints, but a more particular difficulty relating to what deontologists[2] should say about cases of uncertainty. In their key example, a skier is about to cause the death of ten people by causing an avalanche. Jackson and Smith assume that whether or not it is morally permissible (and presumably also—given the possibility of saving the ten—morally required) to kill the skier (this is the only way of saving the ten) depends, according to a typical deontological theory, on whether or not he intends to kill the ten. If so, then he can permissibly be killed in self- (or other-) defense. If not, then it is presumably impermissible to kill him, for presumably a deontological constraint forbids the killing of an innocent person, even in order to save ten innocent people from dying in an avalanche.[3]

The problem, though, is that you do not know for sure whether or not the skier intends to kill the ten. In other words, the probability—*p*—of him being innocent (that is, of him not intending to kill the ten) is lower than 1, but higher than 0. The typical deontological view

---

* For helpful comments on a previous version, we thank Mark Schroeder.

[1] Jackson and Smith, "Absolutist Moral Theories and Uncertainty," this JOURNAL, CII, 6 (June 2006): 267–83.

[2] Jackson and Smith put their challenge as a challenge not to deontology in general, but rather to absolutism (plausibly considered a particular instance of deontology). We explain why reading their challenge as a challenge to deontology in general is more interesting below, in footnote 5.

[3] We are not going to quibble over the details here. In the context of the theory of self-defense, it is highly controversial whether innocent aggressors can be justifiably killed. But nothing in Jackson and Smith's argument—or in our reply to it—depends on the details of the example of the target deontological theory.

0022-362X/08/0505/259–72

Jackson and Smith argue against forbids killing the skier to save the ten if $p=1$, and permits (and perhaps also requires) to kill the skier if $p=0$ But what does such a theory have to say about other values of p? The view according to which the constraint applies in all its force for any $p>0$ is highly implausible, for this view entails—in practical circumstances, where certainty is almost never present—the most extreme form of pacifism The view according to which the constraint does not apply unless $p=1$ is a highly implausible understanding of deontology, for this view is—in practical circumstances, where certainty is almost never preset—extensionally equivalent with typical forms of consequentialism (without certainty, the constraint will just never apply) What else, then, can the deontological view say about cases with intermediate p?

This is, of course, already a theoretically important question But given that in the real world we hardly ever have certainty in such matters, this is also a practical problem A deontological theory that has nothing to say about cases with $1>p>0$ really has nothing to say about practically all real-life cases The severity of the problem thus cannot be overstated

If you are not yet convinced, think about another example, one that is—unlike Jackson and Smith's skier example—entirely practical, and to which we will return below The example of criminal punishment Retributivists about criminal punishment—typically if not always deontologists—believe that there is a deontological constraint against punishing (or "punishing") the innocent, and so that at least sometimes it is wrong to "punish" an innocent person even when doing so promotes the good But what do such theories have to say about the required burden of proof in criminal trials? If we are to require a certainty burden of proof, we will have a highly implausible theory of criminal punishment, according to which punishment is—in practical circumstances, where certainty is almost never present—never justified If we say the deontological constraint against "punishing" the innocent only applies where there is no chance at all of guilt, we effectively take all content out of this deontological constraint So what *should* a retributivist say about the appropriate burden of proof in criminal trials?[4]

---

[4] Here is one possible reply to this question "Retributivism (although not necessarily retributivists) is mute on how high standards of proof ought to be, "—Jeffrey Reiman and Ernest van der Haag, "On the Common Saying That It Is Better That Ten Guilty Persons Escape Than That One Innocent Suffer Pro and Con," in Ellen Frankel Paul, Fred D Miller, Jr , and Jeffrey Paul, eds , *Crime, Culpability, and Remedy* (Cambridge Blackwell, 1990), pp 226–48, see pp 242–43

I. THE THRESHOLD VIEW

Jackson and Smith consider several possible deontological replies, and reject them all. We will not address all of them in this reply. Rather, we are going to limit ourselves to just what seems to us the initially most promising deontological attitude toward uncertainty, and Jackson and Smith's most interesting argument against it. We do more to place this argument in the context of Jackson and Smith's paper in a footnote.[5]

One of the possible solutions Jackson and Smith consider is the one they call "the threshold version of absolutism" (275).[6] We take this to be the obvious thing for the deontologist to say. In the skier example, for instance, there is a certain threshold value for p, such that when p (the probability that the skier does not intend the deaths of the ten) is over that threshold value the deontological constraint applies in all its force, whereas when the value of p is below that threshold, the constraint does not apply at all. Let us call this threshold value "moral certainty." Our deontologist says, then, that if we know to a moral certainty that the skier intends the deaths of the ten, it is morally permissible (and perhaps also required) to kill him in order to save them, but if our (justified) confidence in the skier's murderous intentions does not rise to the level of moral certainty, we ought not to kill him even to save the ten [7]

---

[5] Throughout, Jackson and Smith make the mistake of modeling deontology (or absolutism—we return to the distinction between the two shortly) in very consequentialist-friendly ways. For instance, they consider understanding deontology as (just about) a consequentialist theory that places infinite negative value on constraint-violations, or just very large finite value on such violations. But this will not do, of course, because deontologists do not have to play the consequentialist value-game at all, they are, after all, deontologists.

Also, Jackson and Smith choose to argue just against absolutist moral theories—theories, that is, according to which there are some constraints it is *never* justified to violate. But this restriction reflects a surprising—and disappointing—choice, for hardly any (secular) contemporary deontologist is an absolutist. Contemporary deontologists are typically "moderate deontologists," deontologists who believe that deontological constraints come with thresholds, so that sometimes it is impermissible to violate a constraint in order to promote the good, but if *enough* good (or bad) is at stake, a constraint may justifiably be infringed. By restricting their target theories to just absolutist theories, Jackson and Smith lose just about all initially plausible targets. And indeed, some of Jackson and Smith's arguments do not work against moderate deontological views (in particular, their argument against the "Big Bad Number" view (273–75) essentially relies on the target view being absolutist and not moderately deontological).

But some of what Jackson and Smith say can be abstracted from these worrying problems. The problem mentioned in the previous section is, of course, a serious cause of concern for any deontologist, and the problem they raise for the threshold view—the problem discussed throughout most of this reply—is also generalizable in this way.

[6] It is perhaps worth saying that the thresholds relevant here are different from the ones mentioned in footnote 5. Here these are probability-thresholds, there, they are thresholds regarding how much good (or bad) can justify infringing a deontological constraint.

Applied to the example of criminal punishment The analogous threshold version of retributivism holds that we are only allowed to punish someone if we are confident to the point of moral certainty that she is guilty If our confidence level is lower than moral certainty, we should not punish, even if "punishing" promotes the good Perhaps the traditional burden of proof in criminal trials—beyond a reasonable doubt—is meant to capture something like this view

II JACKSON AND SMITH'S ARGUMENT AGAINST THE THRESHOLD VERSION

Enter Jackson and Smith again, this time with the two-skier example Assume two symmetrical skiers, *X* and *Y*, each in a position identical to that of the skier in the one-skier example with which we started each, that is, about to cause an avalanche that will kill ten (different) people And assume that we know to a moral certainty that *X* intends the deaths of "his" ten, that we know to a moral certainty that *Y* intends the deaths of "his" ten, but—as is surely possible[8]—that we do not know to a moral certainty that *both* intend the deaths of the relevant ten It then follows from the threshold version just described that we should (or at least that it is permissible to) kill *X*, that we should (or at least that it is permissible to) kill *Y*, but that it is impermissible to kill *X* and *Y*—for we do not know to a moral certainty that both intend to kill the (respective) ten, so the deontological constraint against killing the innocent applies

---

[7] The term "moral certainty" has a substantive part to play neither in Jackson and Smith's official statement of the problem, nor in our suggested solution All that matters is that there is some threshold as described in the text We use the term "moral certainty" only to simplify the wording later on But note that Jackson and Smith seem to be using it in a way that renders true the judgment that it is morally permissible to kill the skier whenever we lack moral certainty that he is innocent (that is, that he does not intend to kill the ten) We, on the other hand, use it to render true the inverse judgment namely, that it is morally impermissible to kill the skier unless we know to a moral certainty that he intends to kill the ten We think that our use of the term is more in line with commonsensical intuitions as well as deontological tradition And seeing that nothing hinges on this point here—both Jackson and Smith's argument and our reply can be stated without using the term "moral certainty" at all—there can be no objection to this choice of words on our part

[8] Suppose that moral certainty lies at a probability of 95% And assume that the probability of *X* being innocent (that is, not intending the death of the ten) is 4%, and that the probability of *Y* being innocent is 4% The probability that at least one of them is innocent is then roughly 8% (1–(96* 96) In such a case, we know to a moral certainty that *X* intends the death of the ten, we know to a moral certainty that so does *Y*, and we do not know to a moral certainty that both do

So the threshold version violates, at the very least, Agglomeration—the requirement that if one ought to Φ and one ought to Ψ then one ought to (Φ-and-Ψ). And as if this is not bad enough, the threshold version may be committed here to a highly implausible moral dilemma. For there is no way in which a conscientious agent can both, say, shoot X, shoot Y, and avoid shooting-X-and-Y. So the threshold version has highly implausible implications, and should therefore be rejected.

III. A WAY OUT

In order to explain our way out, it may be useful to start with a simple example. Assume, then, a deontological theory according to which all and only deontological constraints are grounded in rights of individuals. According to such a deontological theory, for instance, the deontological constraint against killing innocent people is grounded in the right innocent people have not to be killed. And in order to avoid the accusation of having nothing to say in the real world, where $1>p>0$, assume a threshold version of such a deontological theory, so that we are not allowed to kill a person (even in order to maximize the good) if we do not know to a moral certainty that she is liable to lethal self-defense (because, for instance, impermissibly attacking others).

How will such a deontological theory fare in the face of Jackson and Smith's two-skier example? Unproblematically, it seems to us. X has a right not to be killed when innocent, but we know to a moral certainty that he intends to kill the ten, and so it is permissible—indeed, given the other features of the example, probably morally required—to shoot X. Similarly for Y. How about shooting-X-and-Y? Well, X-and-Y—presumably the mereological sum of X and Y—is not a right-bearing creature (perhaps rights can be had by entities other than individual persons, like maybe a family, or a nation, but surely to attribute a right to a mereological sum of persons is a category mistake). So there is no deontological constraint that directly requires that we not shoot X-and-Y. So the question whether we know to a moral certainty that both intend the deaths of their respective ten is simply morally irrelevant. Of course, there is no harm in aggregating duties. If we ought to shoot X, and we ought to shoot Y, then we ought in this case to shoot X and Y, and if we want, we can put this point by saying that we ought to perform the complex action of shooting-X-and-Y. But this way of putting things should not stop us from seeing that this complex action is normatively epiphenomenal. Its normative status is completely determined by the normative status of the more simple actions of which it is composed. So according to the kind of theory sketched here, in the two-skier example we ought to shoot *X*, and we ought to shoot *Y*, and (in a normatively epiphenomenal sense) we ought to shoot *X*-and-*Y*, and this result is not problematic, even though we do not know to moral certainty that both *X* and *Y* are attempting murder.

Or consider a sort-of-contractualist deontological moral theory, according to which an action is impermissible (under a deontological constraint) iff someone can reasonably object to its performance.[9] According to such a view, so long as we know to a moral certainty that *X* intends the deaths of the ten, *X* cannot reasonably object to us shooting him, and so we are allowed to shoot him. Similarly, of course, for *Y*. How about shooting *X*-and-*Y*? Well, we know that *X* cannot reasonably object to us so acting, and the same goes for *Y*. Who else *could* object? If *X*-and-*Y* were a morally significant entity, perhaps *X*-and-*Y* could object. But, of course, *X*-and-*Y* is *not* a morally significant entity. So no one can reasonably object to the action of us shooting *X*-and-*Y*.[10] And again, this is simply because shooting *X*-and-*Y* is normatively epiphenomenal on the actions of shooting *X* and of shooting *Y*.

What these examples show, we think, is that Jackson and Smith's argument can be rejected in the following way. The normative status of the complex action shooting-*X*-and-*Y* is completely determined by the normative status of the more basic actions shooting-*X* and shooting-*Y*. Though we can talk about the complex action as well, it is normatively epiphenomenal, and so no deontological constraint applies to it, and so the question of uncertainty at this level does not even arise. When we shoot both, there is no particular person whom we kill while lacking moral certainty of his murderous intentions.[11] And this is sufficient to show that our action does not violate the relevant deontological constraint.

---

[9] This is *very* loosely based on T. M. Scanlon's contractualism. See Scanlon, *What We Owe to Each Other* (Cambridge: Harvard, 1998).

[10] This point is an analogue of the much-discussed issue of aggregation in Scanlon's theory. In that theory one can only reasonably reject principles on individualist grounds. See, for instance, Sophia Reibetanz, "Contractualism and Aggregation," *Ethics*, CVIII (1998): 296–311, F. M. Kamm, "Aggregation and Two Moral Methods," *Utilitas*, XVII (2005): 1–23.

[11] At one point Jackson and Smith put their target theory thus:

> R'': "it is absolutely forbidden to form the intention to kill someone you believe with degree of belief p^* or above to be innocent" (280).

But, contrary to their suggestion, and in line with the thought in the text, R'' is not exposed to the two-skier problem. If a potential shooter follows R'' in the two-skiers case, there is no particular someone whom she believes sufficiently confidently to be innocent. Of course, she is sufficiently confident that at least one of *X*-and-*Y* is innocent. But this is not in violation of R'', as *X*-and-*Y* is not a "someone."

Jackson and Smith consider and quickly reject something like this suggestion (278) They say that the threshold version cannot deal with the two-skier example by saying that only the status of each action counts, because then whether or not we are allowed to shoot *X*-and-*Y* will depend on whether we are using two bullets (one for *X* and one for *Y*, in which case each shooting is allowed, and so so is the complex action) or just one powerful bullet to kill both (in which case *this* is the basic action, and it is wrong, because it violates a deontological constraint) It is important to notice that this reply does not apply to the line of thought sketched earlier We do not claim that what matters is the number of actions performed We claim, rather, that what matters is the nature and grounds of the relevant deontological constraint Under the assumption (of which more shortly) that the relevant deontological constraint has something to do with *X*'s moral status, or with *Y*'s moral status, and that the mereological sum *X*-and-*Y* has no similar moral status, our solution holds So what do we say about the one-powerful-bullet-for-both-skiers example? Jackson and Smith are right in insisting that there cannot be a morally significant difference between killing both *X* and *Y* with one bullet and killing both, each with his own bullet And we already know that we are allowed to kill both with two bullets (because we are allowed to kill each, and because there is no entity *X*-and-*Y* which is a right-bearer, or can object to actions, or some such) So it is also morally permissible to kill both with one powerful bullet This action is permissible, because it does not violate any deontological constraint there is no one person whom we kill even though we lack moral certainty of his intending to kill others

The thing to say, then, about the two-skier example is that—at least for what are arguably the most plausible deontological theories (more on this restriction shortly)—what matters are actions towards *X* and towards *Y* So long as there is no one person whom we kill when lacking moral certainty that they intend the deaths of the ten, the probability that at least one of the relevant group of persons is innocent is not relevant to the applicability of the relevant deontological constraint[12]

---

[12] Jackson and Smith (276–77, fn 10) draw an interesting analogy between their problem for deontology and one version of the lottery paradox In this version, a threshold view of (perhaps justified) belief is seen not to be able to adequately deal with a case of a conjunction of many propositions, each above the threshold, the conjunction itself can still be below it And then a threshold view of (justified) belief arguably leads to inconsistency

Interestingly, we do not think that an analogous reply to the one in the text works for the case of belief For the reply in the text relies on the fact that not any conjunction

IV. INTENDING AND FORESEEING

But how can this be so? Isn't it still true that on our suggested way out your action—say, shooting *X* and *Y* with one powerful bullet—will amount to an intentional killing of someone whose innocence is not ruled out to a moral certainty? And isn't this precisely what the deontological constraint was supposed to rule impermissible?

To see this worry more clearly, think of criminal punishment again. Our suggested solution, when applied to the criminal punishment case, reads something like this: If we "punish" an innocent person (or one in whose guilt we lack moral certainty), this person is entitled to complain, to object to us doing so, her right has been violated. But if we make sure never to punish someone unless we know to a moral certainty that they are guilty, no one is entitled to complain on account of us not knowing to a moral certainty that *all* those we consider punishing are guilty. The moral permissibility of the general punitive policy is just the by-product of the moral permissibility of each of its (normatively significant) components.

But this, of course, raises the following worry: Suppose that beyond-a-reasonable-doubt denotes moral certainty, the threshold distinguishing between where the deontological constraint applies in all its normative force and where it does not. Then the policy of punishing people when their guilt has been proved beyond a reasonable doubt is—according to our analysis here—morally permissible. But with sufficiently many criminal trials, this policy is virtually bound to lead to the "punishment" of innocents. Indeed, the only way to avoid such cases is, it seems, to avoid punishing altogether. But if we do punish—say, after proving guilt beyond a reasonable doubt—the situation is not merely the analogue of the two-skier example, where we lack moral certainty that they are both guilty. Rather, in the criminal punishment case we know—to a moral certainty, easily—that we *will* punish some innocent people. But then how can our suggested way out possibly work in the criminal punishment case? Surely, if we know that a criminal punishment policy will lead to the punishment of the innocent, and if there is a deontological constraint against punishing the innocent, we should not endorse this policy?

---

of normatively significant factors or classifications is itself an independently normatively significant factor or classification (shooting-*X* is normatively significant, shooting-*Y* also is, shooting-*X*-and-*Y* is not, except insofar as some significance follows from the simple actions of which it is composed). But precisely this claim does not work for beliefs. The conjunction of any two propositions is itself a proposition, and one about which we can most sensibly ask whether we should believe it.

We suggest that the deontologist rely here on the distinction between intending and foreseeing (or on some closely related distinction). If we decide to punish someone we know to be innocent (or just don't know to a moral certainty that she is guilty), we *intend* her "punishment" (say, as a means for deterrence). And so we violate the relevant deontological constraint. But if we adopt a policy that—we know—will result in the punishment of innocents, their punishment is not intended but merely *foreseen*.

To see how this helps, think of other deontological constraints. Deontologists who believe in a constraint against harming people typically do not believe in a constraint against acting in ways that *foreseeably* harm people. Of course, that an action foreseeably harms someone is a morally relevant feature of that action, but it does not render it a violation of a deontological constraint against harming people as usually understood. To be a violation, the action has to *intentionally* harm someone. Similarly, then, for the case of punishment. The fact that implementing just about any criminal punishment policy foreseeably leads to the punishing of innocents is morally relevant. But this alone does not suffice for a violation of the constraint against punishing the innocent. What would be necessary for this is a case of intentionally punishing an innocent person. But implementing a criminal punishment policy that is bound to lead to the punishment of some innocents does not amount to intentionally punishing an innocent person (certainly not so long as we require proof to a moral certainty—beyond a reasonable doubt—of guilt as a precondition for punishment).[13]

Getting back to the question at the beginning of this section, then: that you cannot rule out to a moral certainty that your action in the two-skier case (shooting both with one bullet) will amount to an intentional killing of an innocent person is something you foresee, not something you intend. The deontological constraint does not apply.[14]

---

[13] For the suggestion that it is a crucial part of justifying punishment that the foreseen cases of punishing innocents are merely foreseen and not intended, see Alan Wertheimer, "Punishing the Innocent—Unintentionally," *Inquiry*, xx (1977): 45–65, at 53–54.

[14] There is nothing unique here in treating intentions as themselves foreseen and not intended. Discussions of the paradox of deontology and related issues often invoke scenarios where you foresee that (intentional) violations of a deontological constraint will follow your refusing to violate it now.

We cannot discuss here in full some possible complications. For instance, perhaps the deontologist can help herself to the distinction between an ex-ante and an ex-post perspective on things. In terms of an entitlement to complain, in the punishment example no one is entitled to complain ex-ante, though of course the innocent person who is wrongly punished (despite proof beyond reasonable doubt) is entitled to

Of course, the moral significance of the intending-foreseeing distinction is neither uncontroversial nor unproblematic[15] But there is good reason to believe that any deontological theory needs something like this distinction anyway, even regardless of its attitude towards uncertainty[16] So by relying on this distinction here the deontologist does not lose any further points[17] If, in other words, you do not believe the intending-foreseeing distinction can be defended conceptually and normatively, most likely you were not a deontologist even before reading Jackson and Smith

V HOW GENERAL IS OUR WAY OUT?

So this is how we suggest to save the threshold version of the deontological theory from Jackson and Smith's argument The things that really count are the violations of individual constraints, or the violations of individual rights, and while aggregative and future violations are foreseeable, they are not intended, and so do not already constitute a violation now[18]

But it is important to note that this way out does not apply to all deontological constraints, and so nor is it available to all deontological theories The second part of this way out—relying on the intending-foreseeing distinction, or on some distinction very close to it—does seem to be one all deontologists can help themselves to

---

complain ex-post Armed with this distinction, perhaps the deontologist no longer needs also the intending-foreseeing distinction to get the desired results here Nevertheless, we use the intending-foreseeing distinction in the text, for the following reasons First, though some other distinction may do the work here, it seems like there are going to be cases where nothing but the intending-foreseeing distinction will do (Consider, for instance, the distinction between someone employing a criminal justice policy merely foreseeing the conviction of some innocents as a side-effect, with someone intending such wrong conviction as a means for making the system more efficient) Second, as we say in the text, the intending-foreseeing distinction (or some distinction very close to it) is one deontologists probably need anyway, regardless of the discussion of uncertainty And third, we suspect that the ex-ante–ex-post distinction derives its normative weight—if it has such normative weight—from something very much like the intending-foreseeing distinction

[15] For discussion and references, see David Enoch, "Intending, Foreseeing and the State," *Legal Theory*, XIII (2007) 69–99

[16] *Ibid*, see the appendix

[17] Jackson and Smith consider intentions in the context of an attempt to amend the target theory by saying that what is morally forbidden is not in the first place actions but rather intentions They rightly reject such a way of attempting to deal with uncertainty But their reasons for doing so do not apply to the way intentions are used in the text here (as a part of employing the intending-foreseeing distinction)

[18] And notice that this line of defense applies to absolutist theories—those Jackson and Smith explicitly argue against—just as it does to more reasonable deontological theories

(again, not because such distinctions are unproblematic, but rather because deontologists need them anyway). But the first part of this defense is not as universally available. Let us elaborate.

What our solution relies on is the assumption that the relevant deontological constraints are grounded in individualistic patient-based considerations. In other words, we assumed—as is most clear in the examples of deontological theories we used in section III—that the relevant deontological constraints are grounded in thoughts about what individual people are entitled not to have done to them. Without this assumption, we no longer have a way of establishing our main claim in reply to Jackson and Smith, namely that—in the two-skier example—what is of moral significance are only the actions against each skier, not the complex action against both. But not all deontological theories offer this kind of support for their deontological constraints. Think, for example, about a deontological theory according to which we ought not to lie (even to promote the good). And assume that the rationale offered for this constraint has nothing to do with the rights or interests of the one lied to, but rather with the wholeness of the soul of the potential liar. Now construct an analogue of the two-skier example. Say, I am pretty sure of two statements *X* and *Y* that they are true; about each, I am confident to a moral certainty that they are true, but about their conjunction I am not. Am I allowed to assert these statements? Here too, as in Jackson and Smith's original example, it seems like it is permissible to make statement *X*, it is permissible to make statement *Y*, but it is impermissible to make both. But now our line of defense no longer applies. If the rationale of the constraint against lying has nothing to do with the one lied to, then the statement *X*-and-*Y* is a perfectly legitimate candidate for moral relevance, and the moral theory has to give an independent answer as to the permissibility of stating it.

Or consider a patient-based view that nevertheless allows for unproblematic aggregation. According to one example of such a view, if you impose a 0.5 risk on me of being (unjustifiably) killed, I have, as it were, half a reasonable objection at hand, and if there are two people in my situation, then together we have a full reasonable objection, and so the relevant action is wrong. According to such a view it is quite possible that *X*-and-*Y* have available to them together a reasonable objection that neither has on his own, and then it is no longer true that the moral status of shooting-*X*-and-*Y* is normatively epiphenomenal, and then our way out fails, and this deontological theory is again vulnerable to Jackson and Smith's argument. So in order to take advantage of our reply to Jackson and Smith the deontological theory has to be not only patient-based, but also individualistic, that is, it must disallow the kind of aggregation above [19] And notice that the kind of aggregation it must rule out is subtly different from the kind of aggregation more commonly discussed The issue of aggregation is usually brought up when there is no uncertainty, but rather (certain) relatively small harms to a number of individuals, and the question is how to weigh this against a (certain) significantly greater harm to one individual In our case, though, the aggregation is of probabilities, or of expected harm [20] So there is room in logical space for a view that rules out aggregation in the more common sense, but allows for the aggregation of probabilities in the specified way Such a view too, however, cannot utilize our solution to the problem raised by Jackson and Smith (and is arguably implausible anyway[21])

Let us not overstate this point For our solution to apply, it is not necessary to assume any *specific* deontological theory What is required, though, is that the relevant deontological theory be one that grounds constraints in individual patients (for instance, in what rights people have, or in what they are—individually—entitled to object to, or something of this sort)

So Jackson and Smith's argument retains its force against some possible deontological theories (at least, our argument in this paper doesn't show otherwise) But individualistic patient-based deontological theories are off the hook And seeing that these are arguably the more plausible deontological theories anyway, this result is not without interest Indeed, Jackson and Smith's argument can be seen—given our reply—as an argument for the claim that if you are going to be a threshold-deontologist, the deontological constraints you believe in had better be grounded in patient-based, individualistic considerations

---

[19] We want to—and think we can—remain silent on the following interesting question Is what is important here the number of individuals whose rights (say) are violated, or the number of rights violated? Assume a constraint against lying If what is important here is the number of people whose rights are violated, then if I contemplate telling two distinct lies—L_1 and L_2—to the same person, the category L_1-and-L_2 is a relevant category (the person lied to can complain about each lie, but also about both) If what matters is the number of violations, then each of L_1 and L_2 is a violation, and L_1-and-L_2 is epiphenomenal Different deontological theories are going to answer this question differently, and their answer will determine whether they can be defended against Jackson and Smith's argument by employing the defense in the text

[20] If all that matters is expected value (see some discussion in the final section, below), then the two kinds of aggregation are effectively one

[21] On such a view, for instance, and given plausible background assumptions, a deontological constraint against punishment would entail that no criminal punishment system (bound as it is to lead to the punishment of some innocents) is morally permissible

VI REMAINING PROBLEMS

With regard to a wide range of deontological theories, then, a threshold version of the view can satisfactorily deal with Jackson and Smith's argument But the problem raised in Jackson and Smith's paper can be seen as a particular instance of the following family of problems How should moral theories—deontological or otherwise—deal with uncertainty? We hope to have more to say about this wider issue on other occasions Here, though, we wish to conclude by hinting at some of these related problems

One rather predictable difficulty—one anticipated and set aside by Jackson and Smith (276)—is another problem for threshold views It is the problem of finding a non-arbitrary way of placing the relevant threshold But we are not convinced this is a genuine problem, and a special one in this context (rather than an instance of the most general problem of vagueness), and so let us not say anything more about it here

But consider another problem Jackson and Smith's paper deals with uncertainty about the nature of one's action or its results But what if our ignorance runs deeper than mere uncertainty? What should we say of cases in which one of the things we do not know is the very distribution of probabilities over the relevant possibilities?[22] Perhaps we can often deal with uncertainty by estimating the expected value of possible actions But what can we do when—because of such deeper ignorance—we cannot even estimate expected values?[23]

And what should we say if the ignorance involved is not of the relevant (nonmoral) facts (like whether the skier intends to kill the ten) but rather of *moral* facts? Suppose I know that the skier is in fact an innocent aggressor And suppose that—having read much of the literature on self-defense—I am just not sure whether killing an innocent aggressor is morally permissible How should a moral theory deal with this kind of uncertainty?[24] On the one hand, both in this case and in the skier case I end up not knowing for sure whether or not I should kill the skier. On the other hand, there does seem to be a morally relevant difference between factual and moral ignorance or uncertainty. But what is it?

---

[22] There are well-known reasons to think that not all cases of such deeper ignorance are somehow reducible to cases of usual probabilistic uncertainty See, for instance, Donald A Gillies, *Philosophical Theories of Probability* (New York Routledge, 2000), pp 37–49

[23] Notice that this is a problem for consequentialists as well as for deontologists This is an important point, because the problem Jackson and Smith focus on is not, of course, a problem for consequentialism at all, and so one may get the impression that consequentialism faces no serious difficulties in dealing with uncertainty The point in the text shows that this is not so

[24] For some discussion, see Ted Lockhart, *Moral Uncertainty and Its Consequences* (New York Oxford, 2000), Jacob Ross, "Rejecting Ethical Deflationism," *Ethics*, CXVI (2006) 742–68 For a discussion of this point in the context of a discussion of moral

Return now to nonsophisticated, probabilistically quantifiable uncertainty about nonmoral facts. Is it really plausible to say that only expected value counts? In *Trolley*, you have to decide whether to divert a runaway trolley from a track on which it is about to kill five, to another on which it will kill one. In *Trolley**, the situation is similar, except the trolley, if not diverted, is expected with probability of 0.5 to kill ten (and with probability of 0.5 not to kill anyone). In *Trolley***, the undiverted trolley will for sure kill 5, but if diverted it will kill 10 with probability of 0.1 (and will kill no one at all with probability of .9). In *Trolley****, there are 5 on the original track, but on the other track there are 10, each with an independent probability of 0.1 of dying if the trolley is diverted onto that track. If only expected value counts in cases of uncertainty, there is no morally relevant difference between Trolley, Trolley*, Trolley**, Trolley***, and numerous other probabilistic versions of Trolley. But is this a plausible claim?

So it is not at all clear how moral theories—deontological and consequentialist alike—should take uncertainty into account. This topic deserves much more research than it has hitherto received. The fact that Jackson and Smith's argument against deontology can be dealt with does not take anything away from the importance of drawing the attention of moral philosophers to this interesting host of issues.

RON ABOODI
ADI BORER
DAVID ENOCH

Hebrew University of Jerusalem

---

responsibility and blameworthiness, see also Gideon Rosen, "Culpability and Ignorance," *Proceedings of the Aristotelian Society*, ciii (2003): 61–84; Alexander A. Guerrero, "Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution," *Philosophical Studies*, cxxxvi (2007): 59–97.

COMMENTS AND CRITICISM

MORAL ABSOLUTISM DEFENDED*

Frank Jackson and Michael Smith argue against absolutist moral theories [1] A moral absolutist holds that all actions of some kind are, without exception, morally prohibited In Jackson and Smith's view, moral absolutism runs aground when there is uncertainty whether an action is of the prohibited kind Taking as representative the absolutist view that intentionally killing an innocent person is morally prohibited, Jackson and Smith raise a problem case

> A skier is heading in a direction you know for sure will trigger an avalanche that will kill ten people You know the only way to save the ten people is for you to shoot him The probability that the skier intends to trigger the avalanche and kill the ten people is $1-p$ (271)

What ought you to do? This question challenges the absolutist to explain how your uncertainty about the skier's intentions affects what you ought to do Jackson and Smith reject every absolutist answer they can think of, however, as I will argue, they have no good reason to reject the following obvious answer if your degree of belief that the skier is innocent is greater than some threshold *t*, you ought to not shoot the skier If your degree of belief that the skier is innocent is less than or equal to *t*, you ought to shoot the skier, thereby saving the lives of the ten people

Jackson and Smith argue that a moral absolutist cannot accept such a threshold Their main complaint relies on a more complicated example Skier *X* is threatening 10 people, as before But now another skier, *Y*, is threatening a different group of 10 people Your degree of belief that at least one of *X* and *Y* is innocent is above the threshold Your degree of belief that *X* is innocent is below the threshold Your degree of belief that *Y* is innocent is below the threshold You have four possible courses of action shoot both *X* and *Y*, shoot *X* but not *Y*, shoot *Y* but not *X*, or shoot neither

---

*Thanks to Jiwei Ci, Chad Hansen, Joe Lau, William Haines, Michael Martin, Timothy O'Leary, Frank Jackson, and an audience at Hong Kong University The author acknowledges support for this research from grant 20062159005 from the University of Hong Kong Seed Funding Programme for Basic Research

[1] Jackson and Smith, "Absolutist Moral Theories and Uncertainty," this JOURNAL, CIII, 6 (June 2006) 267–83

0022-362X/08/0505/273–75

According to Jackson and Smith, this case poses a serious difficulty for the threshold absolutist because, they claim, each course of action is such that in doing it you fail to do everything you ought to do If you shoot *X* and do not shoot *Y* you fail to do something you ought to do—for you ought to shoot *Y* because your degree of belief that *Y* is innocent is below the threshold If you shoot *Y* and do not shoot *X* you fail to do something you ought to do—for you ought to shoot *X* because your degree of belief that *X* is innocent is below the threshold Similarly, if you shoot neither you fail to do two things you ought to do (shoot *X*, and shoot *Y*) And if you shoot both you do something you ought not to do, because your degree of belief that at least one is innocent is above the threshold

There is no difficulty here for the threshold absolutist The threshold absolutist should not accept that whatever you do in this situation you fail to do everything you ought to do The absolutist can agree that shooting both is morally prohibited You ought not to shoot both because your degree of belief that at least one is innocent is too high Beyond that, the threshold view remains silent, none of the other three actions violate the prohibition against intentionally killing an innocent person But what ought you to do? Various ways to answer are open to the threshold absolutist For example, one simple proposal is that you ought to choose, from the remaining three actions, an action which maximizes expected value Then, if shooting *Y* but not *X* has expected value at least as high as any of these available actions, by shooting *Y* but not *X* you have done what you ought to do There is nothing else that you ought to do

Jackson and Smith disagree They think that the absolutist must accept that if you shoot *Y* but not *X*, you fail to do something that you ought to do The thought is that since your degree of belief that *X* is innocent is below the threshold, you ought to shoot *X* Indeed, say Jackson and Smith, "you ought to shoot *X* *independently* of whether or not you shoot *Y*" (278) However, Jackson and Smith are mistaken, the moral absolutist should not agree that you ought to shoot *X* independently of whether or not you shoot *Y* Explaining why is revealing about the nature of the moral absolutist's commitments

Jackson and Smith do not clearly explain why they think the absolutist should hold that you ought to shoot *X* independently of whether or not you shoot *Y* But it appears that Jackson and Smith have the following metaphysical point in mind "we need to understand [the absolutely prohibited kind] *K* as a property of an action as opposed to a relation between an action and available alternatives to that action" (267) This is because "the distinctive feature of the kind of absolutism that we find, for example, in Kant and the Catholic tradition is that the absolutely prohibited kind is independent of the nature of any available alternatives" (268)

The metaphysical point suggests that if you ought not to perform some action because that action is morally prohibited, you ought not to perform that sort of action whatever the alternatives The threshold absolutist can accept this point According to the threshold absolutist, the moral prohibition kicks in above the threshold—in the one-skier case you ought not to shoot the skier if your degree of belief in innocence is above the threshold So, granting the metaphysical point, the threshold absolutist can agree that in the two-skier case, if your degree of belief that X is innocent is *above* the threshold, you ought not to shoot X, independently of whether or not you shoot Y

However, the metaphysical point only applies to morally prohibited actions, it does not apply to actions which are *not* morally prohibited The threshold absolutist can grant the metaphysical point but maintain that what you ought to do when your degree of belief in innocence is *below* the threshold may depend on the available alternatives The threshold absolutist holds that in the one-skier case, if your degree of belief that X is innocent is below the threshold, you ought to shoot X But, even granting the metaphysical point, it does not follow that you ought to shoot X in the two-skier case when your degree of belief in innocence is below the threshold

Jackson and Smith have no good reason to reject threshold absolutism [2]

PATRICK HAWLEY

Hong Kong University

---

[2] There is a decision theoretic way to argue that in the two-skier case you ought to shoot X whether or not you shoot Y Suppose that the expected value of the complex actions in the two-skier case is the sum of the expected value of each subaction It follows that EV(Shoot Y and Not shoot X) $=$ EV(Shoot Y) $+$ EV(Not shoot X), and EV(Shoot Y and Shoot X) $=$ EV(Shoot Y) $+$ EV(Shoot X) Now suppose that since your belief that X is innocent is below the threshold, EV(Shoot X) $>$ EV(Not shoot X) And similarly suppose that EV(Shoot Y) $>$ EV(Not shoot Y) Then, EV(Shoot Y and Not shoot X) $<$ EV(Shoot Y and Shoot X) But, EV(Not shoot Y and Not shoot X) $=$ EV(Not shoot Y) $+$ EV(Not shoot X), and EV(Not shoot Y and Shoot X) $=$ EV (Not shoot Y) $+$ EV(Shoot X) So, EV(Not shoot Y and Not shoot X) $<$ EV(Not shoot Y and Shoot X) Therefore, neither (Not shoot Y and Not shoot X) nor (Shoot Y and Not shoot X) maximize expected value Every other available action involves shooting X Since you ought to choose an available action maximizing expected value, you ought to shoot X whether or not you shoot Y

The threshold absolutist has a ready reply Although EV(Shoot Y and Not shoot X) $<$ EV(Shoot Y and Shoot X), it does not follow that (Shoot Y and Not shoot X) does not maximize expected value of the available actions Your degree of belief that at least one of X and Y is innocent is greater than the threshold So, the absolutist can say, (Shoot X and Shoot Y) is not an available action because it is morally prohibited The moral prohibition makes (Shoot X and Shoot Y) unavailable for rational deliberation

IN MEMORIAM: MARY MOTHERSILL

The editors and trustees of the *Journal of Philosophy* note with sadness the passing of their long-time colleague and fellow editor Mary Mothersill, who died after a long illness on January 22, 2008, at the age of 84. She was born in Edmonton, Canada, and received her B.A. in English from the University of Toronto (1944), her Masters in Philosophy from Radcliffe College (1945), and her Ph.D. in Philosophy from Harvard University (1954). After a series of appointments at Vassar, Columbia, the University of Connecticut, the University of Michigan, the University of Chicago, and City College, she joined the philosophy faculty at Barnard College in 1963, where she remained until her retirement in 1993. For 28 years she was Chair of the Barnard Philosophy Department, and during her tenure she trained many young women who went on to careers in academic philosophy and made a significant mark on philosophy in North America.

In the early part of her career Mary published papers on an astonishing range of topics, from metaethics, moral knowledge, and the nature of art and criticism, to death, feminism, and pornography. In 1984 she published the work she is best known for, *Beauty Restored*, which quickly became, and has remained, a classic text of twentieth-century aesthetic theory. It is a bold, systematic, and elegant book that, very much against the trend of its time, offered a general account of the central problems of aesthetic judgment, arguing that a judgment of beauty can be a genuine judgment despite the absence of laws or principles of taste. In the course of developing her account, she advanced our understanding of such topics as the nature of metaphor, the nature of aesthetic properties and aesthetic value, and the conceptual differences between aesthetics and morality. *Beauty Restored* is a model of clarity and precision that synthesizes philosophical analysis with a dialogue with thinkers from the past, especially Kant and Hume, while remaining grounded in a lifetime of varied aesthetic experience.

Mary was richly, if belatedly, honored for her philosophical achievements. In 1986 she was a Visiting Fellow at Wolfson College, Oxford, and later she was honored as Alfred North Whitehead lecturer at Harvard. In 1998 she was elected President of the Eastern Division of the American Philosophical Association, and in 2003 she was named a Fellow of the American Academy of Arts and Sciences.

Mary's interests were not limited to academic philosophy. She was an avid gardener, for example, who over the course of years transformed her property in Bridgehampton, New York, into a work of art that incorporated sophisticated design and plants from all over the world. In addition to landscape gardening, Mary loved music, especially Bach. She was a long-time board member of the Choral Society of the Hamptons, and she accumulated an impressive collection of recordings, at both her New York City apartment and her beloved home on Long Island. For many years Mary was vice-chair of the National Emergency Civil Liberties Committee, a group founded in 1951 to respond to McCarthyism from a perspective left of the ACLU. During the 1960s and 1970s she was a vocal and active opponent of the War in Vietnam and of University policies that punished students for taking part in campus protests.

With her death the faculty of Barnard College and Columbia University lost an irreplaceable colleague, friend, and source of inspiration. Moreover, as the testimony of countless former students at her memorial service made clear, we also lost a generous and devoted teacher who left an indelible mark on a large number of those she taught and mentored over the past forty-five years.

FREDERICK NEUHOUSER for the *Journal of Philosophy*

The MIT Press

The Odd One In

On Comedy

Alenka Zupančič

"Arguing that our current 'feel good' society has cheapened the value of comedy, Alenka Zupančič brilliantly restores the genre's subversive edge Full of delightful surprises and profound observation, *The Odd One In* is itself odd in the best sense unique, without peer" — Joan Copjec, author of *Imagine There's No Woman*

Short Circuits series, edited by Slavoj Žizek
240 pp, 3 illus, $19 95 paper

Austere Realism

Contextual Semantics Meets Minimal Ontology

Terence E Horgan and Matjaž Potrč

"Though Horgan and Potrc's ontological conclusions are radical, their reasoning is impeccable and comes with a good dose of reflection on the best way of choosing an ontology This book is an important contribution to the growing literature in metametaphysics" — Josh Parsons, Philosophy Department, Otago University

A Bradford Book
Representation and Mind series
232 pp, 1 illus, $32 cloth

Scenario Visualization

An Evolutionary Account of Creative Problem Solving

Robert Arp

"*Scenario Visualization* is a challenging volume that will be appreciated by scholars and is written with a clarity that will also please laypersons interested in learning about cutting-edge thinking on human cognitive evolution" — Todd K Shackelford, Department of Psychology, Florida Atlantic University

A Bradford Book • 208 pp, 10 illus, $30 cloth

Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling

Economic Theory and Cognitive Science

Don Ross, Carla Sharp, Rudy E Vuchinich, and David Spurrett

"*Midbrain Mutiny* gathers together all current, significant research and theory on the psychology and neurology and economics of gambling In this truly excellent book the authors organize that material, reconcile apparent differences, and make the compelling argument that we are on the verge of a significant breakthrough in treatment of disordered gambling" — Howard Rachlin, Psychology Department, State University of New York, Stony Brook

A Bradford Book
Representation and Mind series
328 pp, 6 illus, $38 cloth

The Case for Qualia

edited by Edmond Wright

Philosophical and scientific defenses of Indirect Realism and counterarguments to the attacks of qualiaphobes

A Bradford Book 384 pp, 2 illus, $36 paper

ZONE BOOKS

NOW IN PAPER

Echolalias

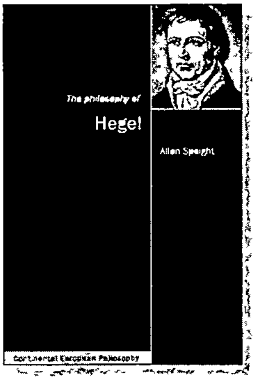
On the Forgetting of Language

Daniel Heller-Roazen

A far-reaching philosophical investigation into the persistence and disappearance of speech, in individuals and in linguistic communities

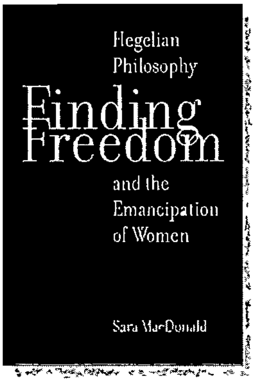
Distributed for Zone Books
288 pp, $21 95 paper

The MIT Press To order call 800-405-1619 • **http //mitpress mit edu**



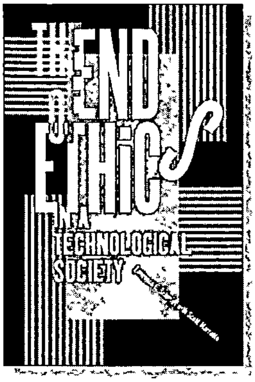
The Philosophy of Hegel
Allen Speight
978 0-7735-3408-7 $22 95 paper | 978-0-7735-3407-0 $80 00 cloth

"An outstanding book It demonstrates a wide and deep knowledge of Hegel and modern Hegel scholarship and an evident enthusiasm for the study of Hegel's ideas Speight is a sure-footed interpreter of Hegel's texts and his selection of particular topics cannot be faulted It is a pleasure to read "
–Dudley Knowles, University of Glasgow



Finding Freedom
Hegel's Philosophy and the Emancipation of Women
Sara MacDonald
978-0-7735-3375-2 $65 00 cloth

"Clear, well-written, and solidly researched *Finding Freedom* puts forward an original thesis concerning Hegel's view of women and the family "
–Paul Franco, Bowdoin College

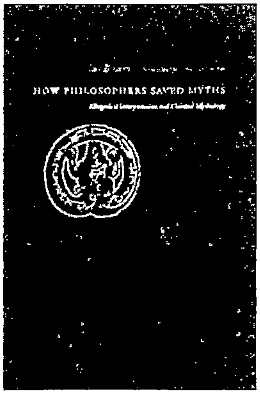


The End of Ethics in a Technological Society
Lawrence E Schmidt with Scott Marratto
978-0-7735-3336-3 $27 95 paper | 978-0-7735-3335-6 $80 00 cloth

"The principle thesis, contemporary moral relativism can lead to moral nihilism, is a very important insight that has been largely overlooked to date, but an idea that needs to be understood and broadly debated in society as a whole "
–Margaret Somerville, director, McGill Centre for Medicine, Ethics, and Law

McGILL-QUEEN'S UNIVERSITY PRESS | www mqup ca

philosophy
from Chicago

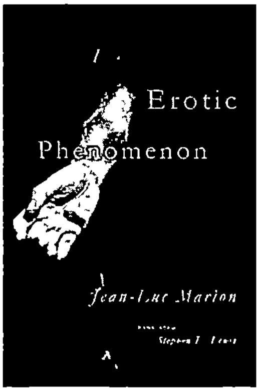


Now in paperback

HOW PHILOSOPHERS SAVED MYTHS
Allegorical Interpretation and
Classical Mythology
Luc Brisson
Translated by Catherine Tihanyi

"In his brilliant and lively book, Brisson follows the history of the link between myth and philosophy all the way up to the Renaissance This wonderful book confirms Brisson's status as one of the major authorities in the field of classical antiquity"—*Choice*

PAPER $19 00



THE EROTIC PHENOMENON
Jean-Luc Marion
Translated by Stephen E Lewis

"Marion's avowed topic is the erotic phenomenon, and his method is phenomenology He is a master of that method, and the result is an analysis of erotic love of unparalleled precision and depth The depiction he gives of the erotic phenomenon is fundamentally convincing, and readers will find their own loves illuminated and questioned"—*Commonweal*

PAPER $19 00

CSLI
PUBLICATIONS

New from CSLI

MEANING, INTENTIONS, AND ARGUMENTATION
Kepa Korta and Joana Garmendia

Meaning, Intentions, and Argumentation brings together the work of leading contemporary scholars—from the old disciplines of philosophy and rhetoric to the newest thinking on semantics and pragmatics—to illuminate crucial aspects of meaning, communication, argumentation, and persuasion

PAPER $32 50

THE UNIVERSITY OF CHICAGO PRESS • www press uchicago edu

New from Stanford University Press

Anatomy of the Passions

FRANÇOIS DELAPORTE,
Translated by
SUSAN EMANUEL,
Edited and with a foreword by
TODD MEYERS

Through the pioneering work of Duchenne de Boulogne, Delaporte provides a remarkable philosophical and historical examination of expressive physiology during the mid-19th century, and considers the science of emotion as a means of revealing inner life—thoughts, feelings—upon the surface of the face

Cultural Memory in the Present
$24 95 paper $65 00 cloth

Silencing the Demon's Advocate
The Strategy of Descartes' Meditations

RONALD RUBIN

"Much has been written on the *Meditations* in the last thirty years, so it is difficult to say anything new, but Professor Rubin's lively and analytically written book succeeds in doing just that Especially noteworthy are his fresh readings of the method of doubt, the *cogito*, and the Cartesian Circle "
—Larry Nolan,
Marquette University

$50 00 cloth

Stanford
University Press

800.621.2736 www.sup.org

You can now subscribe to the *Journal of Philosophy* using

PayPal

To learn more about this easy new option, click on the PayPal icon at **www.jphil.org.**

THE
ART
OF
LIVING

Philosophy is the great untapped resource of our generation – from Plato to Bertrand Russell, philosophers have provided insights about the matters that concern us all, but today their ideas are often ignored in favour of those provided by self-help works. **The Art of Living Series** aims to bring philosophical insights to a wider audience. Taking its lead from the concerns of the ancient Greek philosophers, the series asks the question "How should we live?" Authors draw on their personal reflections as well as their philosophical training to write books that enrich, stimulate, and challenge readers' ideas about how to live their lives.

Illness by Havi Carel
978-1-84465-152-8

Pets by Erica Fudge
978-1-84465-156-6

Clothes by John Harvey
978-1-84465-150-4

Deception by Ziyad Marar
978-1-84465-151-1

Sport by Colin McGinn
978-1-84465-148-1

Fame by Mark Rowlands
978-1-84465-157-3

Work by Lars Svendsen
978-1-84465-154-2

Hunger by Raymond Tallis
978-1-84465-155-9

Wellbeing by Mark Vernon
978-1-84465-153-5

$18.95 paper
160pp

McGILL-QUEEN'S UNIVERSITY PRESS | www.mqup.ca

CAMBRIDGE

New and Noteworthy Titles

The Apology Ritual
A Philosophical Theory of Punishment
Christopher Bennett
$90.00 Hb 978-0-521-88072-5 224 pp

Mathematics, Models, and Modality
Selected Philosophical Essays
John P. Burgess
$99.00 Hb 978-0-521-88034-3 316 pp

The Cambridge Introduction to Michel Foucault
Lisa Downing
Cambridge Introductions to Literature
$70.00 Hb 978-0-521-86443-5
$19.99 Pb 978-0-521-68299-2 160 pp

The Cambridge Introduction to Walter Benjamin
David Ferris
Cambridge Introductions to Literature
$70.00 Hb 978-0-521-86458-9
$19.99 Pb 978-0-521-68308-1 184 pp

The Cambridge Companion to Galen
Edited by R. J. Hankinson
Cambridge Companions to Philosophy
$85.00 Hb 978-0-521-81954-1
$29.99 Pb 978-0-521-52558-9 472 pp

The Cambridge Companion to Hume
2nd Edition
Edited by David Fate Norton and Jacqueline Taylor
Cambridge Companions to Philosophy
$95.00 Hb 978-0-521-85986-8
$32.99 Pb 978-0-521-67734-9 400 pp

The Cambridge Companion to Jung
2nd Edition
Edited by Polly Young-Eisendrath and Terence Dawson
$90.00 Hb 978-0-521-86599-9
$29.99 Pb 978-0-521-68500-9 378 pp

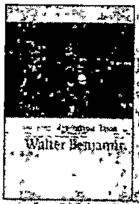
Friedrich Nietzsche and the Politics of History
Christian J. Emden
Ideas in Context
$99.00 Hb 978-0-521-88056-5 412 pp

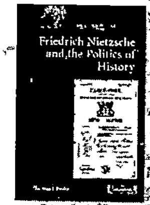
The Morality of Embryo Use
Louis M. Guenin
$90.00 Hb 978-0-521-87269-0
$29.99 Pb 978-0-521-69427-8 288 pp

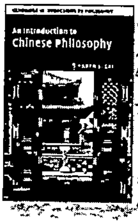
Liberal Beginnings
Making a Republic for the Moderns
Andreas Kalyvas and Ira Katznelson
$65.00 Hb 978-0-521-89946-8
$19.99 Pb 978-0-521-72828-7 200 pp

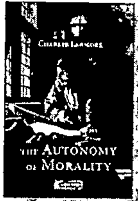
An Introduction to Chinese Philosophy
Karyn L. Lai
Cambridge Introductions to Philosophy
$90.00 Hb 978-0-521-84646-2
$34.99 Pb 978-0-521-60892-3 328 pp

The Autonomy of Morality
Charles Larmore
$85.00 Hb 978-0-521-88913-1
$24.99 Pb 978-0-521-71782-3 290 pp









Prices subject to change

www.cambridge.org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

from Cambridge

Aversive Democracy
Inheritance and Originality in the Democratic Tradition
Aletta J. Norval
$90.00 Hb 978-0 521-87842-5
$29.99 Pb 978-0 521-70268-3 248 pp

Paradoxes of Political Ethics
From Dirty Hands to the Invisible Hand
John M. Parrish
$95.00 Hb 978 0 521-87355 0 296 pp

Mathematical Thought and Its Objects
Charles Parsons
$95.00 Hb 978 0 521-45279-3 398 pp

Barbarism and Religion
Barbarians, Savages and Empires
Volume 4
J. G. A. Pocock
$34.99 Pb 978 0-521-72101-1 384 pp

Leadership Ethics
An Introduction
Terry L. Price
$80.00 Hb 978 0 521-87583-7
$24.99 Pb 978-0-521-69911-2 254 pp

Against Throne and Altar
Machiavelli and Political Theory Under the English Republic
Paul A. Rahe
$90.00 Hb 978-0 521-88390-0 432 pp

Emerson: Political Writings
Edited by Kenneth S. Sacks
Cambridge Texts in the History of Political Thought
$70.00 Hb 978 0 521-88369 6
$23.99 Pb 978-0-521-71002-2 280 pp

Kant's Groundwork of the Metaphysics of Morals
An Introduction
Sally Sedgwick
Cambridge Introductions to Key Philosophical Texts
$90.00 Hb 978-0-521-84345-4
$29.99 Pb 978 0 521-60416-1 224 pp

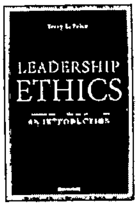
Plato and Theodoret
The Christian Appropriation of Platonic Philosophy and the Hellenic Intellectual Resistance
Niketas Siniossoglou
Cambridge Classical Studies
$95.00 Hb 978-0 521-88073-2 280 pp

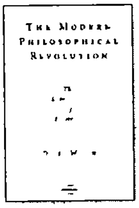
Knowledge and Politics in Plato's *Theaetetus*
Paul Stern
$90.00 Hb 978 0-521-88429-7 328 pp

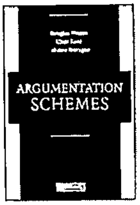
The Modern Philosophical Revolution
The Luminosity of Existence
David Walsh
$90.00 Hb 978-0-521 89895 9
$29.99 Pb 978 0-521-72763-1 520 pp

Informal Logic
A Pragmatic Approach
2nd Edition
Douglas Walton
$85.00 Hb 978-0-521-88617-8
$29.99 Pb 978 0-521-71380-1 350 pp

Argumentation Schemes
Douglas Walton, Christopher Reed, and Fabrizio Macagno
$90.00 Hb 978-0 521 89790 7
$32.99 Pb 978 0 521-72374 9 436 pp







Prices subject to change

www.cambridge.org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS

New From
OXFORD

On Human Rights
JAMES GRIFFIN

What is a human right? How can we tell whether a proposed human right really is one? How do we establish the content of particular human rights, and how do we resolve conflicts between them? These are pressing questions for philosophers, political theorists, jurisprudents, international lawyers, and activists James Griffin offers answers in his powerful new theory of the foundations of human rights

2008 296 pp $50 00

How Terrorism Is Wrong
Morality and Political Violence
VIRGINIA HELD

How Terrorism is Wrong collects articles by Virginia Held that offer a moral assessment of various forms of political violence, with terrorism the focus of much of the discussion Held explores questions of who has legitimate authority to engage in justifiable uses of violence, whether groups can be responsible for ethnic violence, and how the media should cover terrorism

2008 224 pp $45 00

Henry of Harclay
Ordinary Questions, I-XIV
Edited by MARK G HENNINGER

This is the first complete edition of the later work of the medieval philosopher and theologian Henry of Harclay Because of the clarity of Harclay's thought and style, now mirrored in the English translation, the *Quaestiones Ordinariae* are an ideal way to introduce students to key problems in medieval philosophy, as well as to enable scholars to deepen their knowledge of the debates of this period

(British Academy)

2008 400 pp $115 00

Henry of Harclay
Ordinary Questions, XV-XXIX
Edited by MARK G HENNINGER

This volume completes the first full critical edition of the later work of the medieval philosopher and theologian Henry of Harclay, together with an English translation prepared in collaboration with Raymond Edwards

(British Academy)

2008 400 pp $115 00

The Necessity of Theater
The Art of Watching and Being Watched
PAUL WOODRUFF

"Paul Woodruff has written a provocative call to arms that dares to reconsider the very nature of the theater and convincingly explains why live drama has to be an essential part of contemporary society A modern *Poetics*—exciting, innovative and incredibly inspiring Anyone who seriously cares about the future of our theater must read this book"—Peter Meineck, Artistic Director, Aquila Theatre Company & New York University

2008 272 pp $27 95

Everyday Aesthetics
YURIKO SAITO

Yuriko Saito presents a ground-breaking book on aspects of our everyday experience that have been neglected by modern Western aesthetic theories By analyzing a wide range of examples from our aesthetic interactions with nature, the environment, everyday objects, and Japanese culture, she not only shows how exploring everyday aesthetics can enrich the content of our aesthetic discourse, but also reveals the surprisingly powerful influence that our everyday aesthetic tastes and judgments exert on the state of the world and our quality of life

2008 256 pp $49 50

OXFORD
UNIVERSITY PRESS

Prices are subject to change and apply only in the US To order, please call 1-800-451-7556 In Canada, call 1-800-387-8020 Visit our website at **www oup com/us**

12.1.09
UNIVERSITY
OF
CALCUTTA

THE JOURNAL OF PHILOSOPHY

VOLUME CV, NUMBER 6
JUNE 2008

page
277 *The Unity of Consciousness and the Split-Brain Syndrome* Tim Bayne
301 *Are Analytic Philosophers Shallow and Stupid?* Eugene Mills
320 *Guilt by Statistical Association Revisiting the Prosecutor's Fallacy and the Interrogator's Fallacy* Carmen de Macedo

***Published by* The Journal of Philosophy, Inc.**

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J E WOODBRIDGE AND WENDELL T BUSH

Purpose To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines

Editors Bernard Berofsky, Akeel Bilgrami, John Collins, Arthur C Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Wolfgang Mann, Frederick Neuhouser, Christopher Peacocke, Carol Rovane, Achille C Varzi *Consulting Editors* David Albert, James T Higginbotham, Robert May, Charles D Parsons, Wilfried Sieg *Managing Editor* John Smylie

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc *President,* Arthur C Danto, *Vice President,* Akeel Bilgrami, *Secretary,* Michael J Mooney, *Treasurer,* Liaquat Ahamed, *Other Trustees* Lee Bollinger, Leigh S Cauman, Barbara Gimbel, Kent Greenawalt, Lynn Nesbit, Daniel Shapiro

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027 FAX (212) 932-3721

You may also visit our website at wwwjournalofphilosophy org

THE JOURNAL OF PHILOSOPHY

2008

SUBSCRIPTIONS (12 issues)

Individuals	$45 00
Libraries and Institutions	$100 00
Libraries and Institutions (print and online)	$180 00
Students, retired/unemployed philosophers	$20 00
Postage outside the U S	$15 00

Payments only in U S currency on a U S bank All back volumes and separate issues available back to 1904 Please inquire for advertising rates, ad space is limited, so ad reservations are required

Published monthly as of January 1977, typeset by Dartmouth Journal Services, Waterbury, VT, and printed by The Sheridan Press, Hanover, PA

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027 (212) 866-1742

All materials published in *The Journal of Philosophy* are copyrighted by the *Journal* and are protected by U S copyright law All rights reserved For more information about our copyright policies, please consult our January issue or contact us

The Journal of Philosophy (ISSN 0022-362X) is published monthly by The Journal of Philosophy, Inc , MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027 Periodicals postage paid at New York, NY, and other mailing offices

POSTMASTER Please send address changes to *The Journal of Philosophy* at MC4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027

© Copyright 2008 by the Journal of Philosophy, Inc ISSN 0022-362X

THE JOURNAL OF PHILOSOPHY
VOLUME CV, NO 6, JUNE 2008

THE UNITY OF CONSCIOUSNESS AND THE SPLIT-BRAIN SYNDROME*

According to conventional wisdom, the split-brain syndrome puts paid to the thesis that consciousness is necessarily unified The aim of this paper is to challenge that view I argue both that disunity models of the split-brain are highly problematic, and that there is much to recommend a model of the split-brain—the switch model—according to which split-brain patients retain a fully unified consciousness at all times Although the task of examining the unity of consciousness through the lens of the split-brain syndrome is not a new one—such projects date back to Thomas Nagel's seminal paper[1] on the topic—the time is ripe for a re-evaluation of the issues

I THE SPLIT-BRAIN SYNDROME

First performed on humans in the late 1930s, the split-brain procedure involves severing the corpus callosum in order to prevent epileptic seizures spreading from one hemisphere to another The original version of the procedure, known as a commissurotomy, involved severing a number of interhemispheric tracts (such as the anterior commissure, the hippocampal commissure, and the massa intermedia of the thalamus) in addition to the corpus callosum In later versions of the procedure, known as a callosotomy, only the corpus callosum is sectioned The differences between these two patient groups are not pronounced, and I will refer to both commissurotomy and callosotomy patients as "split-brain patients "[2]

---

*In writing this paper I have benefited greatly from comments by Andrew Brook, Michael Corballis, David Chalmers, Barry Dainton, Rami Gabriel, Uriah Kriegel, David Milner, Nishat Nishat, David Rosenthal, Colwyn Trevarthen, Eran Zaidel, and especially Jerre Levy

[1] Nagel, "Brain Bisection and the Unity of Consciousness," *Synthese*, XXII (1971) 396–413

[2] The main division appears to be between patients who have had the anterior portion of their corpus callosum sectioned and those in whom only the posterior portion of

0022-362X/08/0506/277–300

The split-brain procedure has surprisingly little impact on cognitive function in everyday life[3] Split-brain patients can drive, hold down jobs, and carry out routine day to day tasks Early researchers remarked on their "social ordinariness," and were baffled by their inability to detect any cognitive impairments arising from the operation[4] However, subsequent research has revealed a complex array of deficits—and the occasional benefit—in the split-brain[5] It is this research that gives rise to the view that split-brain patients have a disunified consciousness

In a typical split-brain experiment, two stimuli are presented to the patient in such a way that one will be processed by the left hemisphere and the other by the right hemisphere For example, the word 'key-ring' might be projected such that 'key' is restricted to the patient's left visual field (LVF) and 'ring' is restricted to the patient's right visual field (RVF) The contralateral structure of the visual system ensures that stimuli projected to the LVF are processed in the right hemisphere and vice-versa Other perceptual systems can be studied in a similar manner For example, tactile perception is examined by asking the patient to compare an object presented to the right hand with one presented to the left

Such studies have revealed two kinds of disunities in the split-brain *behavioral disunities* and *representational disunities* Behavioral disunities

---

the corpus callosum has been cut The former tend to exhibit the classic split-brain syndrome, while the latter show only minimal dissociations

[3] See Dahlia W Zaidel, "A View of the World from a Split-Brain Perspective," in E M R Critchley, ed , *The Neurological Boundaries of Reality* (Northvale, NJ Aronson, 1995), pp 161–74, S M Fergusen et al, "Neuropsychiatric Observation on Behavioral Consequences of Corpus Callosum Section for Seizure Control," in A G Reeves, ed , *Epilepsy and the Corpus Callosum* (New York Plenum, 1985), pp 501–14 But see also Victor Mark, "Conflicting Communicative Behavior in a Split-Brain Patient Support for Dual Consciousness," in S R Hameroff et al, eds , *Towards a Science of Consciousness* (Cambridge MIT, 1996), pp 189–96

[4] Andrew J Akelaitis, "A Study of Gnosis, Praxis and Language following Section of the Corpus Callosum and Anterior Commissure," *Journal of Neurosurgery*, I (1944) 94–102 For discussion, see Joseph E Bogen, "The Callosal Syndromes," in Kenneth M Heilman and Edward Valenstein, eds , *Clinical Neuropsychology* (New York Oxford, 1993), pp 337–407

[5] Useful reviews of the split-brain literature can be found in Michael S Gazzaniga, "Cerebral Specialization and Interhemispheric Communication Does the Corpus Callosum Enable the Human Condition?" *Brain*, CXXIII (2000) 1293–336, Sally P Springer and Georg Deutsch, *Left Brain Right Brain* (New York W H Freeman, 1998, 5th ed), S E Seymour et al , "The Disconnection Syndrome Basic Findings Reaffirmed," *Brain*, CXVII (1994) 105–15, J J Sidtis, "Can Neurological Disconnection Account for Psychiatric Dissociation?" in Jacques M Quen, ed , *Split Minds/Split Brains Historical and Current Perspectives* (New York University Press, 1986), pp 127–48, George Wolford et al , "Split Decisions," in Gazzaniga, ed , *The Cognitive Neurosciences III* (Cambridge MIT, 2004), pp 1189–200, E Zaidel et al , "The Callosal Syndromes," in Heilman and Valenstein, eds , *Clinical Neuropsychology*, pp 347–403

are most striking. When asked to report what she sees the patient in the key-ring experiment will typically say that she sees only the word 'ring', yet, with her left hand, the patient may select a picture of a key and ignore pictures of both a ring and a key-ring. Generally speaking, visual information projected to the RVF cannot be verbally reported, and visual information projected to the LVF is unavailable for behavior involving the right hand. In the tactile modality, the patient cannot describe, or use her right hand to respond to, objects palpitated by her left hand, and objects palpitated by the right hand cannot be reported via left-handed actions.

Representational disunities involve a lack of integration between the contents of the patient's conscious states. These states do not enjoy the inferential promiscuity that conscious states typically enjoy. The patient in the key-ring experiment appears to have representations of the words 'key' and 'ring' without having a representation of the word 'key-ring'. Similarly, a patient might appear to be conscious of the identity of the objects palpitated by each hand but have no conjoint awareness of both objects. As we shall see, the precise nature of behavioral and representational disunities differs from patient to patient, but the foregoing description captures the core features of the split-brain syndrome.

II. THE UNITY OF CONSCIOUSNESS

Whether or not the split-brain syndrome is at odds with the unity of consciousness clearly depends on what it is for consciousness to be unified. There are a number of things that might be meant by "the unity of consciousness," only some of which are called into question by the split-brain syndrome. This section provides a brief overview of the conception of the unity of consciousness with which I will work.[6]

The kind of consciousness in which I am interested is *phenomenal* consciousness. States of phenomenal consciousness are characterized by the fact that there is something it is like to be in them. Typically, we enjoy multiple phenomenal states at a time. I currently have visual experiences associated with seeing these words on a computer screen, auditory experiences of the sounds associated with the café in which I am sitting, a range of bodily sensations, emotional and mood experiences, experiences of agency, and conscious cognitive states of various kinds Perhaps there is no point in time at which my overall phenomenal perspective includes experiences drawn from each of these categories, but there are certainly times at which I enjoy a number of these phenomenal states at once

---

[6] I draw here on Tim Bayne and David J. Chalmers, "What Is the Unity of Consciousness?" in Axel Cleeremans, ed., *The Unity of Consciousness: Binding, Integration and Dissociation* (New York: Oxford, 2003), pp. 23–58; see also Bayne, *The Unity of Consciousness* (New York: Oxford, forthcoming). For other perspectives on the unity of consciousness, see Barry Dainton, *Stream of Consciousness: Unity and Continuity in Conscious Experience* (New York: Routledge, 2000); John R. Searle, "Consciousness," *Annual Review of Neuroscience*, xiii (2000): 557–78; and Michael Tye, *Consciousness and Persons* (Cambridge: MIT, 2003).

Experiences, when they occur simultaneously, do not occur as phenomenal atoms but have a conjoint phenomenology—there is something it is like to have them together, and they are so-had There is something it is like to taste a well-made macchiato, there is something it is like to have a word on the tip of one's tongue, and there is something distinctive that it is like to enjoy these two phenomenal states *together* One can think of phenomenal unity as a relation that phenomenal states have when they are experienced together in this way When one experiences the taste of coffee together with having a word on the tip of one's tongue, one is in a phenomenal state that in some way subsumes both the sensory and cognitive states

We are now in position to say what it is for consciousness to be unified A subject's consciousness is unified exactly when they enjoy a single phenomenal state that subsumes each of their fine-grained phenomenal states This total phenomenal state—what is sometimes referred to as a "phenomenal field"—fully captures what it is like to be the subject of experience Subjects that are not in such a total state do not have a unified consciousness, there is no single thing that it is like to be such a subject

The split-brain data suggests that split-brain patients do not enjoy a unified consciousness, for it appears as though there are times at which the split-brain patient has no total phenomenal state Of course, this pressure can be relieved by individuating subjects of experience in phenomenal terms, one might say that where an organism has two total phenomenal states, it also has (or supports) two subjects of experience Those tempted by this move will claim that any evidence for thinking that split-brain patient have a disunified consciousness is evidence not that such patients are subjects with a disunified consciousness, but that split-brain patients are not themselves conscious subjects

There is much to be said on behalf of this response but it can be set to one side here, for I will argue that there is no time at which the split-brain *patient* has a disunified consciousness We need not individuate subjects of experience in phenomenal terms in order to "save" the unity of consciousness from the split-brain syndrome At least, so I will argue

III THE CASE FOR PHENOMENAL DISUNITY

Disunity models of the split-brain regard the split-brain patient as having simultaneous, but phenomenally disunified, experiences I will examine two lines of argument for this view of the split-brain, each of which proceeds on the assumption that in the key-ring experiment (and others like it) the patient has representations of the stimuli that are both *simultaneous* and *conscious* I will revisit the assumption of simultaneity in section IV, but for now let us proceed by accepting it What about the assumption that the split-brain patient's mental states are *conscious*?

Early treatments of the split-brain sometimes presented the minor (typically right) hemisphere as a zombie, conscious neither of its environment nor of its own behavior This view was sometimes defended on the grounds that consciousness in the split-brain (and perhaps more generally) is restricted to the language-generating left hemisphere [7] This defense cannot be supported For one thing, a number of split-brain patients have some capacity for both LH and RH speech production [8] More importantly, we should not make language production—or, for that matter, language comprehension—a precondition on the possession of consciousness, such a principle would, implausibly, remove pre-linguistic children and aphasics from the realm of the conscious Even the claim that creatures capable of producing verbal reports must be able to report the contents of each of their conscious states is unacceptably demanding

One might attempt to defend the zombie model by assimilating right-hemisphere guided behavior to other instances of nonconscious behavior, such as on-line dorsal stream motor control in the visual system, or the high-level automaticity effects as studied in social psychology [9] This line of argument is more potent but it too fails to convince, for the right-hemisphere is capable of carrying out tasks that far surpass in cognitive complexity those that are elsewhere attributed to "zombie systems." Summarizing an experiment testing the right-hemisphere based abilities of split-brain patients (with left hemisphere language), R. W. Sperry and others write,[10] "The overall level of the right hemisphere's ability to identity test items and also the quality of the accompanying emotional and evaluative responses were of the same order approximately as those obtained from the right visual field and left hemisphere" (*ibid.*, p. 163). In short, there is good reason to think that each hemisphere in the split-brain can support consciousness—the question is whether the conscious states that they support are unified.

---

[7] See John C Eccles, *The Understanding of the Brain* (New York McGraw-Hill, 1973), D M MacKay, "Cerebral Organization and the Conscious Control of Action," in Eccles, ed , *Brain and Conscious Experience* (New York Springer, 1966), pp 422–45

[8] See K. Baynes et al , "The Emergence of the Capacity of a Disconnected Right Hemisphere to Name LVF Stimuli Implications for Functional Plasticity," *Neuropsychologia*, XXXI (1995) 1225–42, Gazzaniga et al , "Collaboration between the Hemispheres of a Callosotomy Patient Emerging Right Hemisphere Speech and the Left Hemisphere Interpreter," *Brain*, LXXXVIII (1996) 1255–62, Zaidel et al , "The Callosal Syndromes "

[9] For discussion of unconscious behavioral control as revealed by cognitive neuropsychology, see A D Milner and M.A. Goodale, *The Visual Brain in Action* (New York Oxford, 1995/2006), and Andy Clark, "Visual Experience and Motor Action Are the Bonds Too Tight?" *Philosophical Review*, CX (2001) 495–519, for social psychology, see John A Bargh and Melissa Ferguson, "Beyond Behaviorism On the Automaticity of Higher Mental Processes," *Psychological Bulletin*, CXXVI, 6 (2000) 925–45

The first argument for phenomenal disunity appeals to the connection between phenomenal unity and representational content. Consider a split-brain patient (*S*) in the key-ring experiment. *S* has two experiences, one of which represents the word 'key' and one of which represents the word 'ring', but it seems clear that *S* does not have an experience of the word 'key-ring'. But—so the argument goes—any subject with phenomenally unified experiences of the words 'key' and 'ring' must also have an experience of the word 'key-ring'. Since *S* has no such experience, we should conclude that *S*'s experiences of 'key' and 'ring' are not phenomenally unified. Call this the closure argument.

As stated, the closure argument is unconvincing. We must distinguish experiences with the content <'key-ring'> from experiences with the content <'key' & 'ring'>. To see a stimulus as the word 'key-ring' goes beyond seeing it as containing the words 'key' and 'ring'. But this objection is far from fatal, for there is no more reason to suppose that *S* has an experience with the content <'key' & 'ring'> than there is to suppose that *S* has an experience with the content <'key-ring'>. After all, *S* shows no indication—either by verbal report or manual behavior—of having seen the words 'key' and 'ring'.

Of course, absence of evidence is not always evidence of absence. Perhaps, as David Chalmers and I have suggested, *S* has a <'key' & 'ring'> experience whose content is not available for high-level cognitive consumption due to the presence of processing bottlenecks.[11] In support of this proposal we invoked George Sperling's experiments on the reportability of information in brief visual displays. Although the interpretation of Sperling's data is contested, arguably it provides some reason to think that subjects can be conscious of more than they can report.[12]

---

[10] Sperry et al., "Self-Recognition and Social Awareness in the Deconnected Minor Hemisphere," *Neuropsychologia*, xvii (1979): 153–66.

[11] Bayne and Chalmers, "What Is the Unity of Consciousness?"

But there are problems with this proposal. First, subjects in the Sperling case have a kind of meta-access to their putatively "unreportable" contents, for they report *that* they were aware of more than they could directly report. Split-brain patients produce no such reports; unlike Sperling's subjects, they do not say that they had fleeting "unreportable" experiences. Furthermore, one cannot appeal to processing bottlenecks to explain why the patient cannot use the contents of her representation of 'key' in ways in which she can use her representation of 'ring' and vice-versa, for it seems unlikely that saying 'key' or picking out a ring with one's left hand presents more of a challenge to consuming systems than does saying 'ring' or picking out a key.

Let us return to the closure argument. The argument can be presented as follows:

(1) *S* has experiences with contents <*A*> and <*B*>.
(2) *S* does not have an experience with content <*A*&*B*>.
(C) Therefore, *S*'s experiences with contents <*A*> and <*B*> are not phenomenally unified.

There is clearly a missing premise here. We can tighten the argument up by appealing to the following principle:

> *Closure*: Necessarily, if a subject (*S*) has an experience with content <*A*> and an experience with content <*B*>, and these experiences are phenomenally unified with each other, then *S* has an experience with content <*A*&*B*>.

Closure is not incontrovertible. One could treat phenomenal unity as a primitive relation that has no implications for the representational relations between the contents of those states that it unifies.[13] Nonetheless, *closure* is appealing. I am inclined to think that experiences are phenomenally unified only when they stand in a certain relation—subsumption—to a single phenomenal state. And if this is right, then there is a plausible line of argument for the closure principle. Suppose that phenomenal states *A* and *B* are unified. In that case, there will be a phenomenal state (*M*) that subsumes *A* and *B*. Now, it seems plausible to suppose that if *M* subsumes *A* and *B* then the content of *M* must entail the contents of *A* and *B* In the case of states with contents <*A*> and <*B*>, the best candidate for a subsuming state will be a state with content <*A*&*B*> In short, *closure*, and with it the closure argument, is highly plausible

---

[12] Sperling, "The Information Available in Brief Visual Presentations," *Psychological Monographs*, LXXIV, 11 (1960): 1–29. See also Ned Block, "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience," *Behavioral and Brain Sciences*, XXX (2007): 481–548.

[13] See Dainton, *Stream of Consciousness*.

A second argument for phenomenal disunity in the split-brain involves an appeal to behavioral unity, or rather, the lack thereof

(1) *S*'s representations of 'key' and 'ring' are not behaviorally unified although the contents of both states are available for high-level consumption, they are not available to *the same* consuming systems
(2) Phenomenal unity entails behavioral unity two experiences cannot be phenomenally unified without being behaviorally unified
(C) So, *S* has simultaneous but phenomenally disunified experiences

The crucial premise would appear to be (2) Does behavioral disunity entail phenomenal disunity?

There is some reason to think that the contents of consciousness need not be globally available for high-level control And if that is right, then it is possible that the contents of phenomenally unified conscious states might be available for different forms of high-level control—that is, they might not be behaviorally unified Let us examine two syndromes in which behavioral unity seems to break down

Children participating in the Dimensional Change Card Sort task are asked to sort a series of cards (for example, red rabbits and blue dogs) into piles according to a certain dimension (for example, color) [14] Having sorted several cards, the children are then told to switch the sorting rule, say, from colors to animals Three year-olds typically fail to switch dimensions when instructed to do so, but they usually respond correctly to questions about what they ought to be doing Their verbal behavior suggests that they are conscious of the post-switch rules, yet the content of this state does not seem to be available to drive their sorting behavior

The Dimensional Change Card Sort task involves cognitive states Of more direct relevance to the interpretation of the split-brain data is evidence that perceptual states are not always globally available for cognitive consumption Subjects in metacontrast experiments are presented with a series of letters in a format designed to "mask" some of the stimuli In one such experiment, subjects were instructed to press one key if they saw the letter 'J' (for example) and another key if they failed to see a 'J' [15] When urged to respond as quickly as possible subjects tended to respond to the occurrence of a target letter in the "blanked" (masked) positions with a fast (and *correct*) press of the "target present" key, only to apologize immediately for having made an error [16] Arguably, these subjects had experiences whose contents were available to some forms of behavioral control (manual button-pressing) but not others (verbal report) [17]

---

[14] See Phillip D Zelazo, "An Age-Related Dissociation between Knowing Rules and Using Them," *Cognitive Development*, xi (1996) 37–63

Nonetheless, even if behavioral disunity does not *entail* phenomenal disunity, it does seem reasonable to regard it as a *good guide* to phenomenal disunity In general, the best explanation of the fact that the contents of (simultaneously) conscious states are not available to the same consuming systems is likely to be that they are not phenomenally unified And, in light of this, the behavioral disunity argument surely has some weight

I have examined two arguments for the claim that split-brain patients are phenomenally disunified the closure argument and the behavioral disunity argument Each argument has considerable merit, and their combined force does much to justify disunity accounts of the split-brain So let us temporarily proceed on that assumption that the split-brain patient is phenomenally disunified The question we must now address is whether the split-brain subject has two separate streams of consciousness or a single, partially unified, stream of consciousness

IV THE TWO-STREAMS MODEL

It is frequently said that split-brain patients have two streams of consciousness [18] However, the two-streams moniker has been applied to a variety of views. Here, I take the two-streams model to hold that at any one time phenomenal states in the split-brain patient can be divided into two sets, *A* and *B*, where states within each set are mutually phenomenally unified but no state within either set is phenomenally unified with any state in the other set.

---

[15] Reported in Alan Allport, "What Concept of Consciousness?" in Anthony J Marcel and E Bisiach, eds , *Consciousness in Contemporary Science* (New York Oxford, 1988), pp 159–82

[16] See also Joel Lachter and Frank H Durgin, "Metacontrast Masking Functions A Question of Speed?" *Journal of Experimental Psychology Human Perception and Performance*, xxv (1999) 936–47, Lachter et al , "Disappearing Percepts Evidence for Retention Failure in Metacontrast Masking," *Visual Cognition*, vii (2000) 269–79, Anthony J Marcel, "Slippage in the Unity of Consciousness," in Gregory R. Bock and Joan Marsh, eds , *Experimental and Theoretical Studies of Consciousness* (New York Wiley, 1993), pp 168–79

[17] These cases do not falsify the claim that conscious content is always globally available for cognitive control One could attempt to rescue the claim that conscious content is always globally available for cognitive control by invoking local performance failures For example, one could say that perceptual content in the metacontrast experiment was report*able* even though not report*ed*. I doubt that this line of response can be sustained, but must leave this issue to one side here

[18] Lawrence H Davis, "Cerebral Hemispheres," *Philosophical Studies*, lxxvii (1997) 207–22, Michael S Gazzaniga and Joseph E LeDoux, *The Integrated Mind* (New York Plenum, 1978), James Moor, "Split-Brains and Atomic Persons," *Philosophy of Science*, xlix (1982) 91–106, Charles E Marks, *Commissurotomy, Consciousness and Unity of Mind* (Cambridge MIT, 1981), Roland Puccetti, "The Case for Mental Duality Evidence from Split-Brain Data and Other Considerations," *Behavioral and Brain Sciences*, iv

The two streams model draws support from standard presentations of the split-brain data of the kind I gave in section I. Such presentations encourage one to conceive of the split-brain operation as bisecting a single global workspace into two (less global) workspaces, one per hemisphere. Unfortunately for the two-streams model, there is rather more integration in the split-brain than this picture would predict. Although the details of inter-hemispheric integration vary from patient to patient, almost all split-brain patients show some degree of inter-hemispheric integration.

Some (partial) split-brain patients are split for visual information but not tactile information.[19] They can integrate tactile information presented to each hand, but cannot integrate visual information presented across the visual mid-line. However, such patients can integrate tactile information presented to each hand with visual information presented in the ipsilateral visual hemi-field. A similar fractionation of integrative abilities can be observed in patients with complete commissurotomies. Although patient N.G. was unable to match patterns presented to her right hand against those presented in her LVF, she could match patterns presented to her left hand against those presented in her RVF.[20] Given that N.G. can (presumably) match patterns presented to her right (left) hand against those presented in the RVF (LVF), she too seems to possess the kind of behavioral integration that is ruled out by the two-streams model. Even within vision standard split-brain subjects are not fully split. Information concerning shape, color, and category typically cannot be integrated between hemifields, but most split-brain patients are to integrate information about the relative motion and size of visual stimuli.[21]

(1981): 93–123; Sperry, "Mental Unity following Surgical Disconnection of the Cerebral Hemispheres," *Harvey Lectures*, LXII (1966–67): 293–323; Tye, *Consciousness and Persons*.

[19] Gazzaniga and Howard Freedman, "Observations on Visual Processes after Posterior Callosal Section," *Neurology*, XXIII (1973): 1126–30.

[20] Eran Zaidel, "Stereognosis in the Chronic Split-Brain: Hemispheric Differences, Ipsilateral Control and Sensory Integration across the Midline," *Neuropsychologia*, XXXVI, 11 (1998): 1033–47.

[21] Colwyn Trevarthen, "Experimental Evidence for a Brainstem Contribution to Visual Perception in Man," *Brain Behavior and Evolution*, III (1970): 338–52; Trevarthen and Sperry, "Perceptual Unity of the Ambient Visual Field in Human Commissurotomy Patients," *Brain*, XCVI (1973): 547–70; see also Michael C. Corballis, "Visual Integration

These data problematize the two-streams model, for they suggest that the split-brain patient has experiences that "straddle" both hemispheres Such bilateral experiences would be phenomenally unified with both right hemisphere and left hemisphere experiences, in opposition to the thought that the patient's two streams of consciousness are sealed off from each other

In response, the two-streams theorist might be tempted to argue that apparently bilateral experience are really confined to one or other of the patient's two hemispheres The *content* of so-called bilateral experiences might involve information drawn from both hemispheres, but—so the proposal goes—the experiences themselves do not bridge the two hemispheres But this proposal faces the following problem Suppose that the two-streamer assigns a supposedly bilateral experience with the content <*A*&*B*> to the patient's right hemisphere (which includes an experience with content <*A*>) Now, does this stream also include an experience with the content <*B*>? If not, it is difficult to see how the experience <*A*&*B*> could have formed But if we allow that *S*'s right hemisphere stream contains an experience with exactly the same content as her left hemisphere stream, then we must endorse the duplication assumption, according to which it is possible for a conscious subject to have, simultaneously, two experiences with exactly the same content I shall argue below that the duplication assumption is problematic

Behavioral integration in experimental contexts also poses challenges for the two-streams view In the typical split-brain experiment, the patient's right-handed behavior accords with her verbal reports and differs from her left-handed behavior However, in some experiments the patient's right-handed behavior accords with her *left-handed* behavior, both of which are at odds with her verbal reports[22]

---

in the Split-Brain," *Neuropsychologia*, xxxiii, 8 (1995) 937–59, Johnson, "Bilateral Visual Cross-integration by Human Forebrain Commissurotomy Subjects", E Zaidel, "Interhemispheric Transfer in the Split-Brain Long-Term Status following Complete Cerebral Commissurotomy," in Richard J Davidson and Kenneth Hugdahl, eds , *Brain Asymmetry* (Cambridge MIT, 1995), pp 491–32 Partial representational integration is made possible by the representational specialization of the corpus callosum—the anterior midbody transfers motor information, the posterior midbody transfers somatosensory information, the isthmus transfers auditory information, and the splenium transfers visual information (De Lacoste et al , "Topography of the Corpus Callosum," *Journal of Neuropathology and Experimental Neurology*, xciv (1985) 578–91, Margaret G Funnell et al , "Insights into the Functional Specificity of the Human Corpus Callosum," *Brain*, cxxiii (2000) 920–26)

[22] See J Levy et al , "Perception of Bilateral Chimeric Figures following Hemispheric Deconnexion," *Brain*, xcv (1972) 61–78 It is unclear to me why the data reported in

Furthermore, some patients (for example, L B) can name LVF stimuli without being able to integrate stimuli between the two visual hemi-fields, whereas others (for example, N G) can integrate stimuli between the two visual hemi-fields but cannot name LVF stimuli[23] Again, these findings suggest that the availability of content to systems of cognitive consumption in the split-brain is a messy and somewhat fragmented affair, rather than one in which there is a clean division between two clearly demarcated workspaces

Of course, there are various ways in which one might attempt to account for inter-hemispheric behavioral integration within the two-streams framework Each hemisphere has some degree of bilateral motor control, and many split-brain patients have at least some right-hemisphere capacity to comprehend and produce language In the light of this, the two-streamer could argue that (say) left-handed responses involve a consuming system that has access to both left and right hemisphere streams of consciousness More perspicuously, perhaps, the two-streamer might regard left-handed responses as involving different consuming systems depending on whether or not they are guided by the right hemisphere or the left hemisphere There are certainly questions to be asked here about just how consuming systems—and cognitive workspaces more generally—ought to be individuated, but I doubt that the two-streams model can be saved by going down this path In fact, the model is likely to look increasingly implausible as the two-streamer is forced to individuate consuming systems in an ad hoc fashion rather than on any principled (or even intuitive) basis

A further objection to the two-streams model concerns everyday integration in the split-brain How could someone with two streams of consciousness exhibit the kind of behavioral unity that split-brain patients demonstrate in their day-to-day lives? Some two-streamers meet this objection by suggesting that split-brain patients have two streams of consciousness only in experimental conditions[24] The main challenge for this contextualist position is to explain how the structure of the patient's consciousness might be altered by the transition between everyday and experimental environments given that phenomenal structure supervenes only on neural structure and neural structure

---

this study depart from those normally reported, but it may have had something to do with the kinds of stimuli Levy's group used

[23] L E Johnson, "Vocal Responses to Left Visual Field Stimuli following Forebrain Commissurotomy," *Neuropsychologia*, xxii (1984) 153–66, Johnson, "Bilateral Visual Cross-Integration by Human Forebrain Commissurotomy Subjects," *Neuropsychologia*, xxii (1984) 167–75

[24] Marks, *Commissurotomy, Consciousness and Unity of Mind*, Tye, *Consciousness and Persons*

seems not to be fundamentally altered by moving between everyday and experimental contexts

Of course, the contextualist could deny that phenomenal structure does supervene on neural structure She might follow Susan Hurley in identifying the unity of consciousness with a "dynamic singularity in the field of causal flows that is centered on but not bounded by a biological organism "[25] But even those attracted to Hurley's vehicle externalism need to explain *how* the transition between everyday and experimental contexts has an impact on the structure of consciousness The contextualist might argue that the cognitive demands (uniquely) imposed by experimental conditions alter the patient's neural dynamics in such a way that the patient's single stream of consciousness bifurcates into two streams, whereas removing those demands reunites these two streams Although we do not know enough about the neural basis of consciousness to rule this proposal out, it does not seem to me to be a promising one After all, high cognitive load does not normally bifurcate the stream of consciousness, so why should it do so in the context of the split-brain syndrome?

Rather than going contextualist, most two-streamers attempt to account for everyday behavioral unity in the split-brain by deploying the *duplication gambit*[26] They hold that the ability of patients to orient at will to salient stimuli allows them to enjoy duplicate experiences—that is, distinct phenomenal states with the same content, one in each conscious stream

The duplication gambit faces three challenges The first concerns its very coherence Can a conscious subject have, at a single time, multiple experiences with the same phenomenal content? Given that many perceptual properties are bilaterally represented, it might seem obvious that the answer to this question must be 'yes' But matters are not so straightforward Phenomenal states are usually individuated in terms of their content (or phenomenal character), subject, and time This tripartite account rules out phenomenal duplicates, for by definition duplicate states have the same content, are had by the same subject of experience, and occur simultaneously So, on the face of things, endorsing the duplication gambit requires rejecting the standard account of experiences.[27] This is not an inconsiderable cost.

---

[25] Hurley, *Consciousness in Action* (Cambridge Harvard, 1998), p 207 See also Hurley, "Action, the Unity of Consciousness, and Vehicle Externalism," in Axel Cleeremans, ed , *The Unity of Consciousness Binding, Integration and Dissociation* (New York Oxford, 2003), pp 78–91

[26] See Davis, "Cerebral Hemispheres", Moor, "Split-Brains and Atomic Persons", and Puccetti, "The Case for Mental Duality"

Some proponents of the duplication gambit might be tempted to deny that apparent duplicates would belong to the same subject of experience. Perhaps in sectioning the corpus callosum we have also created two subjects of experience. I have some sympathy with this response, but I am assuming here that the split-brain patient is but one conscious subject, whether or not he or she lacks a unified consciousness. Note, moreover, that few two-streamers will want to save the duplication gambit in this way, for most are committed to the claim that both streams of consciousness belong to a single subject of experience.

But why insist on the tripartite account of experience? Why should the proponent of the duplication gambit not individuate experiences in (say) neural terms? Perhaps a subject can have multiple tokens of the same experiential type at the same time as long as the two states occur in (supervene on, are grounded in) different neural areas.

Perhaps the most potent motivation for retaining the tripartite account is phenomenological. As representationalists have argued in recent years, it is plausible to hold that we have introspective access only to the content (phenomenal character) of our conscious states.[28] Suppose that you have exactly three experiences, A_1, V_1, and V_2, where A_1 is an auditory experience and V_1 and V_2 are visual experiences with identical content. Further, suppose that A_1 is phenomenally unified with exactly one of your visual experiences (say, V_1). Would you be able to tell, on the basis of introspection alone, that it is V_1 rather than V_2 that is unified with A_1? I think not. You might have introspective access to the fact that you have a V-type experience that is unified with an A-type experience, but you would lack introspective access to the fact that it is V_1 rather than V_2 that is unified with A_1. Indeed, you would have no introspective access to the fact that you have two V-type experiences. Phenomenal unity is an experiential relation. There must be a phenomenal or "what it's like" difference between a state of affairs in which the members of a set of states are phenomenally unified and one in which they are not, and this phenomenal difference must be introspectively accessible, at least in principle. But the possibility of phenomenal duplication would be at odds with this constraint, for it would allow that a state of affairs in which phenomenal unity holds might be introspectively indistinguishable from a state of affairs in which it fails. In short, any account of phenomenal states that allows for phenomenal duplicates jeopardizes our grip on the very notion of phenomenal unity.[29]

---

[27] Someone might argue that we should think of so-called duplicates as parts (or components) of a single token experience rather than distinct tokens of the same experiential type. The problem with this proposal is that it is not clear why distributed neural activity ought to count as the vehicle of a single token experience given that the mental events it gives rise to are neither functionally nor phenomenologically unified.

[28] See, for example, Tye, *Consciousness and Persons*.

Finally, the duplication gambit would be deeply problematic even if we were to recognize the coherence of phenomenal duplication. For one thing, there is no reason to suppose that the *total* content of each of the patient's two streams would be identical. There might be *overlap* in perceptual content between the two streams, but other components of the streams—cognitive, emotional, intentional, and agentive content—would presumably differ, and these differences ought to lead to behavioral disunity. Furthermore, it is not clear that even complete duplication of content between streams would lead to behavioral integration. States with the same content can trigger different behaviors if they are located in different deliberative contexts, and such differences are likely if, as a number of theorists have suggested, the two hemispheres have unique cognitive styles.[30] Stimuli that the left hemisphere regards as salient might not be regarded as salient by the right hemisphere, and vice-versa.

In summary, neither of the strategies deployed by two-streamers to account for everyday behavioral integration in the split-brain is promising. The two-streams model may represent the conventional wisdom on the split-brain, but it does not deserve that status.

V. PARTIAL UNITY

Grappling with the difficulties posed by the data, Nagel suggested that perhaps there is no whole number of minds that split-brain patients enjoy (*op. cit.*, p. 410). One way to develop Nagel's proposal is in the direction of Michael Lockwood's partial unity model, according to which split-brain subjects have simultaneous experiences (E_1, E_2 and E_3) such that E_1 and E_2 are each phenomenally unified with E_3 but not with each other[31] The partial unity model attempts to capture the fact that split-brain subjects show too much unity to be thought of as having two completely separate streams of consciousness, yet too little unity to be ascribed a single fully unified consciousness

---

[29] In effect, I have turned an argument suggested by Hurley on its head (see *Consciousness in Action*, p. 165; see also Hurley, "Action, the Unity of Consciousness, and Vehicle Externalism," p. 74 and 82). Whereas Hurley uses the possibility of phenomenal duplicates to reject the "what it's like" analysis of phenomenal unity ("co-consciousness," in her terminology), I use the "what it's like" analysis of phenomenal unity to reject phenomenal duplicates. See also note 34.

[30] See, for example, Matthew Roser and Michael S. Gazzaniga, "Automatic Brains—Interpretive Minds," *Current Directions in Psychological Science*, XIII, 2 (2004): 56–59.

As we saw in the previous section, the "raw" split-brain data seem to favor the partial unity model over its two-stream rival Far from clustering into two autonomous systems, perception and cognition in the split-brain is inter-hemispherically integrated in various ways (see also section vi) But if the data themselves point to the partial unity model why has it had so few proponents?

I suspect that neglect of the partial unity models is almost entirely due to concerns about its intelligibility consciousness, so the thought goes, *cannot* be partially unified Even Lockwood, to whom we owe the model, admits to having doubts about its coherence[32] I share Lockwood's doubts, but care must be taken in how they are put It is sometimes suggested that the problem with the partial unity model is that we cannot imagine what it would be like to have a partially unified consciousness[33] This, it seems to me, is not the real issue here The phenomenal perspective of a partial unified subject may not be imaginatively accessible (to us, at least), but it would be the height of hubris to suppose that the limits of what we can imagine are the limits of phenomenal possibility[34]

---

[31] To the best of my knowledge, the first explicit presentation of the partial unity model can be found in Lockwood (*Mind, Brain and Quantum*), but the earlier neuropsychological literature contains frequents hints of it See, for example, Sperry, "Mental Phenomena as Causal Determinants in Brain Function," in Gordon Globus et al , eds , *Consciousness and the Brain* (New York Plenum, 1976), pp 163–77, Sperry, "Consciousness, Personal Identity, and the Divided Brain," *Neuropsychologia*, xxii, 6 (1984) 661–73, C Trevarthen, "Analysis of Cerebral Activities That Generate and Regulate Consciousness in Commissurotomy Patients," in S J Diamond and J Graham, eds , *Hemisphere Function in the Human Brain* (New York Wiley, 1974), pp 235–63, Trevarthen, "Functional Relations of Disconnected Hemispheres with the Brain Stem, and with Each Other Monkey and Man," in Marcel Kinsbourne and W L Smith, eds , *Hemispheric Disconnection and Cerebral Function* (Springfield, IL C C Thomas, 1974), pp 187–207

[32] Lockwood, "Issues of Unity and Objectivity," in Christopher Peacocke, ed , *Objectivity, Simulation and the Unity of Consciousness Proceedings of the British Academy*, Volume LXXXIII (New York Oxford, 1994), pp 89–95

[33] Dainton, *Stream of Consciousness*, p 98, Lockwood, *Mind, Brain and the Quantum*, p 92, Peacocke, "Introduction The Issues and Their Further Development," in Peacocke, ed , *Objectivity, Simulation and the Unity of Consciousness*, p xx

[34] Hurley also argues that the (un)imaginability objection to the partial unity model fails, but her argument depends on the rejection of "what it's like" accounts of phenomenal unity ("co-consciousness" in her terminology) This is problematic, for phenomenal unity is defined in what it's like terms Indeed, Hurley herself appears to invoke "what it's likeness" in introducing the notion of co-consciousness *(Consciousness in Action*, p 88*)*

The real objection to the partial unity model concerns not its unimaginability but its *inconceivability*. Contrast the perspective of a partially unified subject with that of a bat. Although the phenomenal perspective of a bat is not imaginatively accessible to us, we have no difficulty conceiving that there is something it is like to be a bat. By contrast, we have difficulty—arguably *great* difficulty—in conceiving that there is something it is like to be partially unified. Arguably, first-person acquaintance with consciousness reveals that simultaneous phenomenal states that are unified with a third (simultaneous) experience *must* be unified with each other.

This line of argument will not convince everyone. Some theorists will not share the intuition that partial unity is impossible, others will put little stock in inconceivability intuitions (at least when they involve consciousness), and even those who both share the intuition that partial unity is impossible and who are prepared to grant it some epistemic weight might regard its force as being outweighed by the empirical considerations in favor of the model. Nonetheless, it seems to me that the apparent inconceivability of partial unity goes *some* way towards undermining its appeal. We ought to be reluctant to accept a model of the split-brain that is of dubious coherence.

It is time to recap. Not only does the two-streams model receive only equivocal support from the experimental data, it has trouble accounting for the everyday behavioral integration of split-brain patients. The partial unity models fares better with respect to the experimental data, and arguably has less difficulty accounting for everyday integration in the split-brain, but its very coherence is questionable. Perhaps we should reconsider the possibility that split-brain patients retain a unified consciousness.

VI. THE SWITCH MODEL

In an important series of split-brain experiments, Levy and collaborators presented chimeric stimuli—that is, stimuli created by conjoining two similar half-stimuli at the vertical midline—to a series of split-brain patients.[35] On some trials patients were instructed to point to the figure that matched the stimulus, whilst on other trials patients were required to name the stimulus.

> For all patients examined, and for tasks including the perception of faces, nonsense shapes, picture of common objects, patterns of Xs and squares, words, word meaning, phonetic images of rhyming pictures, and outline drawings to be matched to colors, patients gave one response on the vast majority of competitive trials. Further, the nonre-

---

[35] Levy, "Perception of Bilateral Chimeric Figures following Hemispheric Deconnexion."

> sponding hemisphere gave no evidence that it had any perception at all. Thus, if the right hemisphere responded there was no indication, by words or facial expression, that the left hemisphere had any argument with the choice made, and, similarly, if the left hemisphere responded, no behavior on the part of the patient suggested a disagreement by the right hemisphere.[36]

In their original report, Levy and co-authors presented a two-streams interpretation of these findings, according to which each hemisphere had a conscious perception of the stimulus presented in the contralateral visual field. But, as Levy subsequently pointed out, this model fails to explain the absence of inter-hemispheric conflict. If the two hemispheres had separate perceptions, why did they not take issue with each other's responses?

In light of this, Levy proposed an alternative—and, I think, superior—account of these experiments and the split-brain in general.[37] I call it the "switch model." As the name suggests, the switch model holds that consciousness in the split-brain switches between the patient's two hemispheres. The hemispheres contribute in succession to the contents of the patient's consciousness, but, for the most part at least, consciousness does not occur in both hemispheres simultaneously. The switch model paints the split-brain patient as suffering from a kind of fluctuating perceptual extinction: when the left hemisphere is activated stimuli in the RVF win the competition for entry into consciousness at the expense of LVF stimuli, and the converse happens when the right hemisphere is activated. In general, inter-hemispheric activation will march in step with changes in the subject's attentional focus. Rapid inter-hemispheric switches will generate the impression that the patient is conscious of much more than she is in fact conscious of—in much the same way, perhaps, that our fluid interaction with the environment generates the impression that we are conscious of more than we are.

From the perspective of the switch model, we can now see that the closure and behavioral disunity arguments go wrong in assuming that the patient is *simultaneously* conscious of 'key' and 'ring'. The patient might be conscious of the word 'key' (due to right hemisphere activation), and she might be conscious of the word 'ring' (due to left hemisphere activation), but she will not be conscious of both 'key' and 'ring' at the same time, even when the two words are presented simultaneously The patient's behavior might suggest that she is simultaneously conscious of both stimuli, but this would be an illusion generated by the rapidity with which her attention switches between hemispheres[38]

---

[36] Levy, "Regulation and Generation of Perception in the Asymmetric Brain," p. 235.

[37] Levy, "Manifestations and Implications of Shifting Hemi-Inattention in Commissurotomy Patients," *Advances in Neurology*, XVIII (1977): 83–92, and Levy, "Regulation and Generation of Perception in the Asymmetric Brain," in Trevarthen, ed., *Brain Circuits and Functions of the Mind: Essays in Honour of Roger W. Sperry* (New York: Cambridge, 1990), pp. 231–48.

In their original report of the chimeric studies, Levy and co-authors ascribed a nonexpressed conscious state to the nonresponding hemisphere on the grounds that the patient could be encouraged to report (or express) its percept simply by changing the response required of the subject[39] However, they came to regard this modulation in responses as facilitating the entry into consciousness of previously unconscious content, rather than merely allowing the patient to access hitherto inaccessible experiences

> lateralization of seeing of human commissurotomy patients to one half of the visual field was influenced by requiring response with one or the other hand, and the side of perception could be switched by interchanging hands In the case of human subjects, the lateralization of perception was, on most occasions, further and more profoundly effected by asking the subject to speak about his experiences, causing him to use the left hemisphere and to shift his vision to favour the right field[40]

In other examples of response-dependent processing, Levy and Trevarthen found that requiring patients to match chimeric stimuli based on their visual appearance favored the LVF (that is, RH processing) whereas instructions to match chimeric stimuli based on their function favored the RVF (that is, LH processing)[41] Arguably, however, consciousness is not shuttled between hemispheres by changes to the patient's responses per se but by the redistribution of attention that such changes bring about[42]

---

[38] E L Teng and Sperry found further evidence of perceptual extinction in the split-brain (Teng and Sperry, "Interhemispheric Interaction during Simultaneous Bilateral Presentation of Letters or Digits in Commissurotomized Patients," *Neuropsychologia,* xi (1973) 131–40) They presented six split-brain patients with a mixture of dot and numeral counting exercises, in which the stimuli were presented either in the LVF or RVF alone, or in the two visual fields simultaneously Patients were able to report stimuli presented to either field on trials involving only one visual hemi-field, but showed massive amounts of extinction on bilateral trials

[39] It is this interpretation that has entered this philosophical literature In an odd irony, Marks (*op cit*, p 47, n 26) rejects the switch model on the grounds that it is at odds with Levy's chimeric experiments

[40] Trevarthen, "Functional Relations of Disconnected Hemispheres with the Brain Stem," p 195

[41] Levy and Trevarthen, "Metacontrol of Hemispheric Function in Human Split-Brain Patients," *Journal of Experimental Psychology Human Perception and Performance,* ii (1976) 299–312

[42] Something akin to inter-hemispheric switching can be elicited in normal subjects, see A D Milner and J J Dunne, "Lateralized Perception of Bilateral Chimaeric Faces by

Levy's chimeric experiments provide reasonably direct evidence in favor of the switch model, but there are additional considerations that lend a degree of background plausibility to it One such consideration concerns the structure of attention in the split-brain In a 1987 review paper Michael Gazzaniga described the attentional system in the split-brain as "largely integrated "[43] Gazzaniga's assessment was perhaps an over-simplification, but there is evidence that many attentional systems remain unified in the split-brain [44] An early study by C Krueter and others[45] concluded that "a maximum effort by one hemisphere does withdraw capacity from the other, an effect which in the absence of the corpus callosum is presumably mediated by a 'capacity distributing system' located in the brain stem" (*ibid*, p 460) More recently, J D Holtzman and Gazzaniga showed that cognitive load in one split-brain hemisphere can interfere with performance in the other hemisphere [46] A J Lambert has argued that there is a single system of selective attention in the split-brain,[47] while

---

Normal Subjects," *Nature*, CCLXVIII, 5616 (1977) 175–56 Milner and Dunne used chimeric stimuli in which the vertical join was hidden by a white strip, the purpose of which was to hinder detection of the incongruity between the two sides of the stimulus At 100 ms exposure normal subjects had great difficulty detecting that the stimuli were chimeric On trials in which no awareness of asymmetry was present, the subjects indicated (either manually or verbally) only one face, which was always perceived as complete Furthermore, Milner and Dunne's subjects manifested response-dependent processing akin to that seen in Levy's experiment, with verbal responses favoring RVF stimuli and left-handed responses favoring LVF stimuli One *could* take this study to show that normal subjects have two streams of consciousness under these experimental conditions, but it seems to me more reasonable to conclude that even in the normal brain visual experience can switch between hemispheres

[43] Gazzaniga, "Perceptual and Attentional Processes following Callosal Section in Humans," *Neuropsychologia*, XXV, 1A (1987) 119–33

[44] See M Arguin et al, "Divided Visuo-Spatial Attention Systems with Total and Anterior Callosotomy," *Neuropsychologia*, XXXVIII (2000) 283–91, J D Holtzman et al, "Spatial Orientation following Commissural Section," in R Parasuraman and D R Davies, eds, *Varieties of Attention* (Orlando Academic, 1984), pp 375–94, S J Luck et al, "Independent Hemispheric Attentional Systems Mediate Visual Search in Split-Brain Patients," *Nature*, CCCXLII, 6249 (1989) 543–45, Luck et al, "Independent Attentional Scanning in the Separated Hemispheres of Split-Brain Patients," *Journal of Cognitive Neuroscience*, VI (1994) 84–91, G R Mangun et al, "Monitoring the Visual World Hemispheric Asymmetries and Subcortical Processes in Attention," *Journal of Cognitive Neuroscience*, VI (1994) 267–75

[45] Kreuter et al, "Are Deconnected Cerebral Hemispheres Independent Channels? A Preliminary Study of the Effect of Unilateral Loading on Bilateral Finger Tapping," *Neuropsychologia*, X (1972) 453–61

[46] Holtzman and Gazzaniga, "Dual Task Interactions Due Exclusively to Limits in Processing Resources," *Science*, CCXVIII (1982) 1325–27

[47] Lambert, "Interhemispheric Interaction in the Split-Brain," *Neuropsychologia*, XXIX, 10 (1991) 941–48, Lambert, "Attentional Interaction in the Split-Brain Evidence from Negative Priming," *Neuropsychologia*, XXXI, 4 (1993) 313–24

H Pashler and co-authors[48] concluded that "even after commissurotomy the left and right hemispheres appear to be incapable of selecting motor actions independently and simultaneously" (*ibid*, p 2383) The degree of attentional integration seen in these studies is not what the two-streams and partial-unity models would lead one to expect, but it is very much in keeping with the switch account

The switch model also acquires indirect support from a wide range of evidence—drawn from studies of coma, the minimally conscious state, anaesthesia and hydranencephelic children—that identifies sub-cortical systems, centered on the thalamus, as playing a crucial role in consciousness[49] These systems are not divided in the split-brain patient, which might explain not only how consciousness could switch between hemispheres but also why split-brain patients, unlike cetaceans, have a single sleep-wake cycle[50]

I conclude by considering two objections to the switch model If, as the switch model asserts, split-brain patients often undergo sudden and radical changes in the contents of their experience, why are they not aware of this? After all, subjects who experience alterations in the contents of consciousness (as in binocular rivalry) are normally aware of those alterations Split-brain patients occasionally report sudden changes in the contents of consciousness, but such comments are rare Why might this be?

One possibility is that inter-hemispheric switches in consciousness might themselves be unusual outside of experimental contexts Perhaps split-brain patients generally get by on one (presumably the left) conscious hemisphere Experimenters often remark that it can be difficult to elicit right-hemisphere controlled behavior in the split-brain Even when the right hemisphere initiates a task the left frequently takes over and attempts to complete it, sometimes to the detriment of the patient's performance[51]

---

[48] Pashler et al, "Sequential Operation of Disconnected Cerebral Hemispheres in Split-Brain Patients," *Neuroreport*, v (1994) 2381–84

[49] Michael T Alkire and Jason Miller, "General Anesthesia and the Neural Correlates of Consciousness," in Steven Laureys, ed, *Progress in Brain Research, Volume 150 The Boundaries of Consciousness* (2005), pp 445–55, Laureys, "The Neural Correlate of (Un)awareness Lessons from the Vegetative State," *Trends in Cognitive Sciences*, ix, 12 (2005) 556–59, Nicholas D Schiff and Fred Plum, "The Role of Arousal and Gating Mechanisms in the Neurology of Impaired Consciousness," *Journal of Clinical Neurophysiology*, xvii (2000) 438–52, B Merker, "Consciousness without a Cerebral Cortex A Challenge for Neuroscience and Medicine," *Behavioral and Brain Sciences*, xxx, 1 (2007) 63–81

[50] L M Mukhametov, "Interhemispheric Asymmetry of the Electroencephalographic Sleep Patterns in Dolphins," *Brain Research*, cxxxiv (1977) 581–84

[51] Sperry, "Lateral Specialization in the Surgically Separated Hemispheres," in Francis O Schmitt and Frederick G Worden, eds, *Neuroscience, 3rd Study Prog* (Cambridge MIT,

But why don't split-brain patients experience phenomenal discontinuity in those contexts in which they *do* undergo an inter-hemispheric switch in consciousness? I lack a full response to this question, but there is some comfort to be had in the fact that disorders of consciousness are often accompanied by "introspective blindness." In their famous study of unilateral neglect, E. Bisiach and C. Luzzati asked patients to imagine themselves standing in Milan's Piazza del Duomo with their back to the cathedral.[52] As predicted, they failed to describe the buildings on the left. But when asked immediately afterwards to describe when they would see if looking at the cathedral from the opposite end of the square, the same patients named the previously neglected buildings and neglected those that they had just mentioned. At no point did the patients attempt to integrate their successive reports, nor did they express any concern about the obvious inconsistency between them. Just as the ability to track perceptual continuity may be impaired by the very damage that causes unilateral neglect, so too the ability to detect changes in the contents of consciousness may be undermined by the very procedure that prevents consciousness from being bilaterally distributed. It is one thing for the contents of one's consciousness to switch, it is another to be conscious *of* switches in conscious content.

A second objection concerns not the truth of the switch model per se but rather the claim that it is consistent with the unity of consciousness. I have presented the switch model in terms of a single stream of consciousness switching between hemispheres, but a critic might claim that the switch model is better described in terms of the possession of two streams of consciousness that are activated only sequentially. Thus, the critic might continue, far from vindicating the claim that consciousness remains unified in the split-brain, the switch model actually undermines it.

This objection returns us to the question of what it is for consciousness to be unified. According to the account formulated in section II, all it takes for a subject to have a unified consciousness at a time is the existence of a single phenomenal state (or phenomenal field) that subsumes each of the subject's experiences at that time. As far as I

1974), pp. 5–19; E. Zaidel and Sperry, "Performance on the Raven's Colored Progressive Matrices Test by Subjects with Cerebral Commisurotomy," *Cortex*, IX (1973): 34–39; R. D. Nebes and Sperry, "Hemispheric Deconnection Syndrome with Cerebral Birth Injury in the Dominant Arm Area," *Neuropsychologia*, IX (1971): 247–59.

[52] Bisiach and Luzzatti, "Unilateral Neglect of Representational Space," *Cortex*, XIV (1978): 129–33.

can see, there is nothing in the critic's redescription of the switch scenario that is at odds with this account

Of course, the critic might reply that even if the switch model is consistent with the unity of consciousness, there is a sense in which it is at odds with the *continuity* of consciousness in order for a subject to have a single stream of consciousness during a particular interval, that subject's experiences must be grounded in a single set of consciousness-generating mechanisms And, so the critic continues, this condition is not met if consciousness alternates between left and right hemispheres

There is something to the claim that the continuity of consciousness requires physical continuity of some kind, although spelling out just what kind of continuity is required is not easy No matter, for there is good reason to think that the switch model is consistent with any plausible continuity requirement I presented the switch model as holding that consciousness in the split-brain switches between hemispheres, but this is true only to a first approximation Two important qualifications must be made First, some forms of conscious content, such as affective content, involve sub-cortical systems that are not separated by the split-brain procedure The mechanisms responsible for such states constitute a form of physical continuity that underlies inter-hemispheric switches Second, we should not think of the cortical mechanisms responsible for the content of consciousness as the mechanisms of consciousness per se Arguably, cortical activity does not generate consciousness under its own steam, but contributes to the contents of consciousness only when integrated with sub-cortical processing Again, these sub-cortical networks can provide any physical continuity that might be required for the continuity of consciousness

VII CONCLUSION

Although few contemporary theories would follow the seventeenth-century anatomist Giovanni Lancisi in identifying the corpus callosum as the seat of the soul, it is widely assumed that splitting the corpus callosum also splits consciousness The burden of this paper has been to undermine this dogma I have attempted to loosen the grip that disunity models of the split-brain have on us, and to present the switch model as a live alternative to them Not only does it do better in accounting for the behavior of split-brain patients in both experimental and everyday contexts, it also avoids the philosophical baggage that accompanies the two-streams and partial unity models

I leave open the question of whether consciousness is necessarily unified Even if consciousness in the split-brain syndrome remains unified, it is possible that the unity of consciousness breaks down in the context of other pathologies of consciousness, and, of course, it is possible that the unity of consciousness might fail in nonhuman animals All that can be said at this stage is that the case against the unity of consciousness remains unproven

TIM BAYNE

St Catherine's College, The University of Oxford

ARE ANALYTIC PHILOSOPHERS SHALLOW AND STUPID?*

The paradox of analysis concludes that

> (*PA*) A conceptual analysis is correct only if it is trivial

Philosophers from Socrates onward have essayed conceptual analyses of knowledge, freedom, truth, goodness, and more The paradox of analysis suggests that these philosophers—"analytic philosophers" on one apt use of that label, a use I here adopt—are shallow and stupid shallow because they stalk triviality, stupid because it so often eludes them Even if you philosophize without analyzing, you should chafe at this dispiriting indictment of colleagues and predecessors

I will offer a solution to the paradox of analysis that grants *PA*, while arguing that analytic philosophers can embrace *PA* without embarrassment Focusing on the nature and upshot of the triviality attaching to successful conceptual analysis, I will argue that accepting *PA* does not commit you to the intellectual sterility of such analysis Analysis can tax our abilities and enrich our understanding even though its result, when it succeeds, is triviality

A philosophical paradox is best expressed as an argument with premises that seem obviously true, reasoning that seems obviously valid, and a conclusion that seems obviously absurd Solution requires rejecting at least one of these seemings while plausibly diagnosing its intuitive appeal Solving the paradox of analysis, then, is not necessarily a matter of explaining how a correct conceptual analysis can be nontrivial, to presuppose otherwise is to prejudge the proper solution to the paradox It is rather a matter of sympathetically diagnosing either the flaw in the argument leading to *PA* (if *PA* is false) or else the mistaken basis for thinking *PA* false (if *PA* is true)

I start, then, with the argument Consider the sentences

(*BT*) A boy is a boy

---

*Ancestors of this paper received useful discussion at the 2000 Bled Conference on Philosophical Analysis, at a colloquium in my own department in 2001, and at the 2002 annual meeting of the Virginia Philosophical Association I am especially indebted for discussion to Anthony Ellis, Robert Redmon, Peter Vallentyne, Wai-hung Wong, and the late Earle Coleman Geoff Goddu served as commentator at the VPA meeting and gave valuable feedback Takashi Yagisawa offered detailed, helpful comments and prevented me from broadcasting my weak grasp of Japanese An anonymous referee for this JOURNAL offered good suggestions for revision I am grateful to all

0022-362X/08/0506/301-19

and

(*BA*) A boy is a male human child[1]

Each of the following claims has strong pre-theoretical appeal

(1) *BT* is trivial
(2) *BA* provides an analysis of the concept expressed by 'boy'
(3) If *BA* provides an analysis of the concept expressed by 'boy', then 'boy' is synonymous with 'male human child'
(4) If 'boy' is synonymous with 'male human child', then *BA* is synonymous with *BT*
(5) Triviality is closed under synonymy (That is, any sentence synonymous with a trivial sentence is trivial)
(6) *BA* is not trivial

On a uniform interpretation of their constituents, these individually plausible sentences form an inconsistent set and thus a paradox To render it argumentative, take (1)–(5) as premises These imply the falsity of (6), but the allure of (6) dignifies its rejection with the name of paradox

Let me forestall a potential misunderstanding *BA* is not *formally* analytical Someone might understand *BA* as a mere particular affirmative sentence *BA* and *BT* should be read instead as synonymous abbreviations of 'Being a boy is being a male human child' and 'Being a boy is being a boy', respectively I use the shorter formulations for brevity's sake

PA waits in the wings Just add that for any correct analysis, analogues of (1)–(5) will hold pre-theoretical appeal and entail that the analysis is trivial Denying *PA* does not commit you to holding that all correct conceptual analyses are nontrivial For most analyses that concern philosophers, though, analogues of (6) also appeal[2]

If philosophers stuck to analyses like *BA*, the charge of shallowness would stick Interesting cases involve interesting concepts—freedom, knowledge, mentality, and so on[3] Why, then, do I focus on a dull case? Because no interesting analysis is uncontroversially correct The paradox of analysis besets correct analyses, so it is best illustrated by a clearly correct case

---

[1] If 'child' connotes humanity, then 'human' is redundant in *BA*. I resist such connotation, so that (for example) foals are children If this usage grates, substitute 'juvenile' for 'child'

[2] The paradox of analysis is due to G E Moore, see "A Reply to My Critics," Paul Arthur Schlipp, ed , in *The Philosophy of G E Moore* (La Salle, IL Open Court, 1968), pp 533–677 I put the paradox in my own terms, but my resolution of it adapts straightforwardly to Moore's version

[3] I neither explain nor defend here the claim that these concepts hold philosophical interest, but I do not assume that this claim is inexplicable or indefensible

To solve the paradox, you must motivate and defend a denial of at least one of (1)–(6), explaining the intuitive appeal of the claim(s) you reject Familiar solutions include denying any of (2)–(5) While I think that all of these fail, I lack room here to canvass and rebut them I offer, at any rate, a rival solution against which more familiar ones must at least be judged I will, however, discuss one familiar solution briefly, because it raises considerations relevant to my own unfamiliar solution, on which I will then focus I reject (6) and by extension accept *PA* I defend this solution by first diagnosing the appeal of (6)—clarifying relevant senses of 'triviality' along the way—and then showing how the truth of *PA* is compatible with the fruitfulness of conceptual analysis

I THE METALINGUISTIC LINE

One proposed solution to the paradox that needs special mention invokes metalinguistic meaning Translation helps to motivate and explain this "metalinguistic line" on the paradox Translate (1)–(6) into Japanese for the benefit of a monolingual Japanese philosopher, Hiroko Call the Japanese translation of (1) 'J_1', of (2) 'J_2', and so on Now, J_2 refers to the English sentence *BA* and mentions the English word 'boy' Hiroko will find no basis in her linguistic competence for believing J_2 Not understanding *BT*, *BA*, 'boy', or 'male human child', J_5 is the only one of J_1–J_6 that she will find plausible Only J_5 mentions no English expression that she does not understand So for her, J_1–J_6 raise no paradox She will recognize that they form an inconsistent set while having no idea which of the statements is false She will nevertheless see no paradox because J_1–J_4 and J_6 have no independent plausibility for her (She might yield to your authority and believe that all are plausible for those who understand them, but this possibility is irrelevant She will not find them plausible herself)

Try, then, to capture the paradox in Japanese by translating all the mentioned expressions, including *BT* and *BA* Call the resulting sentences 'J_1'', 'J_2'', and so on While these sentences will not be synonymous with (1)–(6), they will parallel them and should, it seems, convey the paradox to Hiroko

The Japanese translation of *BT* is

(*JBT*) Otoko no ko wa otoko no ko desu

The Japanese translation of *BA* is

(*JBA*) Otoko no ko wa otoko no ko desu

We have a problem *JBT* = *JBA* Japanese has no single word for 'boy' but translates it as 'otoko no ko' 'otoko' translates 'male human' while 'ko' translates 'child' (The 'no' is a grammatical particle) *JBA* is trivial to Hiroko in just the way that *BT* is trivial to an English-speaker J_6', which says that *JBA* is not trivial, will seem clearly false to Hiroko Similarly for J_2', which says that *JBA* provides an analysis of the concept expressed by 'otoko no ko' Once again, the paradox fails to survive translation

A familiar explanation of this failure of translation attributes partly metalinguistic meaning to those uses of *BA* on which it turns out nontrivial I will focus on just one of the relevant possibilities, which is that part of what is meant by 'boy' is a condition associated with the word 'boy' itself A crude version would say that (on some uses) 'John is a boy' is synonymous with 'John is a thing called "boy"' This is certainly too crude an account of standard English uses of 'boy' It is not clearly mistaken, though, to hold that what many English-speakers mean by 'boy' is *partly* metalinguistic As was pointed out long ago,[4] certain oddities of translation (which would include those involving *BA* and *JBA*) are explicable by finding metalinguistic content in terms that do not wear it on their sleeves

If—to put it compactly if slightly misleadingly—the word 'boy' is part of its own meaning but not part of the meaning of 'male human child', this fact will explain how *BA* can be nontrivial It will also explain why *JBA* does not accurately translate *BA*, since no English word is part of the meaning of 'otoko no ko' The crux of this option is that a person who finds *BA* informative means by 'boy', 'male', 'human', or 'child' something at least partly about one of those very words, and *BA* is nontrivial by virtue of supplying metalinguistic information

The metalinguistic line promises an account of how *BA* can be nontrivial—of how (6) can be true To solve the paradox while retaining (6), though, you must say which of (1)–(5) is false, and why Defenders of the metalinguistic line must reject (2) Considering cases will make this clear

---

[4] By Alonzo Church, "Intensional Isomorphism and Identity of Belief," *Philosophical Studies*, v, 5 (1954) 65–73, and C H Langford, "Moore's Notion of Analysis," in Schlipp, ed , *The Philosophy of G E Moore*, pp 319–42 Versions of the metalinguistic line have also been endorsed by Felicia Ackerman in "Analysis, Language, and Concepts The Second Paradox of Analysis," *Philosophical Perspectives*, iv (1990) 535–43, Richard Fumerton in "The Paradox of Analysis," *Philosophy and Phenomenological Research*, xliii, 4 (1983) 477–97, and Moore (*op cit*) See also Benson Mates, "Synonymity," in Leonard Linsky, ed , *Semantics and the Philosophy of Language* (Urbana Illinois UP, 1952), pp 111–36

Consider a first-grader—call him 'Fred'—who understands 'male', 'boy', and 'child', but has yet to learn 'human' Confronting *BA*, he might well find it linguistically illuminating, by inferring from it the meaning of 'human' based on his knowledge of the meanings of its other constituent words He will not find it "trivial" (I lean for now on a merely intuitive notion of triviality Clarification will come) The illumination gained from encountering *BA* is explained in terms of his coming to believe that

(7) Humans are called 'humans'

Of course, Fred cannot articulate this proposition using the *sentence* (7) before learning what 'human' means (or, given the facts of cognitive and linguistic development, for some time after) But if he has the concept *boy*, then he has the concept *human*, even if he lacks a ready means of expressing it

This point underlies the metalinguistic line's rejection of (2) Fred is originally ignorant of (7) If he encounters and accepts *BA* but fails to make the metalinguistic inference—if he learns no new language—then what he gleans from *BA* will be, if anything, something like

(8) A boy is a male child called 'human'

If he does make the metalinguistic inference, he will glean something like

(9) A boy is a male human child called 'human'

According to the metalinguistic line—insofar as it offers a solution to the paradox—*BA* is nontrivial because it means, to Fred, something like what (8) or (9) means to us In either case, (2) is false Recall that *BA* is shorthand for a form that makes explicit reference to concepts *BA* is understood as synonymous with 'Being a boy is being a male human child' According to the metalinguistic line, we must take Fred, accordingly, as interpreting *BA* as saying (what we would express by)

(8′) Being a boy is being a male child called 'human'

or

(9′) Being a boy is being a male human child called 'human'

It is clear, however, that neither of these provides a correct conceptual analysis of the concept expressed by 'boy', for neither of them is true Being called 'human' is not part of (the concept of) being a boy, and there could be boys who are not called 'human'

UNIVERSITY OF CALC

Fred might instead mean by *BA* what we would mean by

(8″) All boys are male children called 'human'

or

(9″) All boys are male human children called 'human',

where (8″) and (9″) are universal affirmative sentences expressing contingent propositions. In that case, (8″) and (9″) are true, and the metalinguistic line gives a promising account of the informativeness of *BA*. On this interpretation, though, *BA* obviously does not provide a conceptual analysis.

No algorithm determines what metalinguistic content to attribute, given the metalinguistic line, when someone finds *BA* nontrivial. The point is that if English-speakers find a sentence of the form 'An *F* is a *G*' nontrivial, then '*F*' and '*G*' are not synonyms in their mouths—even if they sincerely believe otherwise (and even if '*F*' and '*G*' are synonyms in our mouths). People can use terms nonsynonymously without realizing that they do so. Where putative conceptual analyses are concerned, the most plausible diagnosis of nonsynonymy in such cases is that at least one of the putative synonyms has implicit metalinguistic content.

Take the metalinguistic line, and you can offer a motivated denial of (2) that makes its pre-theoretical appeal intelligible. Many of us use ordinary words with a partially metalinguistic content without realizing it (the argument goes). Since 'boy' and 'male human child' have the same nonmetalinguistic content, it is easy to understand how we might mistakenly accept (2).

The metalinguistic line may capture the understanding of some sentences by some speakers, but it does not solve the paradox of analysis. In denying (2), it says in effect that if something appears to be a correct, informative conceptual analysis, then the appearance must deceive: if the candidate-analysis is true, then it must not be a conceptual analysis. So the metalinguistic line offers no rebuttal to *PA*. It tells us how *BA* can be correct and nontrivial only by interpreting it as something other than a conceptual analysis. It offers no account of how a conceptual analysis might be correct and nontrivial.

Notice, too, that the metalinguistic line succeeds only for speakers relevantly like Fred. It is implausible, though, that we are like Fred with respect to most of the terms involved in analyses of philosophical interest. We are not like him with respect to *BA*, and so the metalinguistic line offers only an attenuated solution even in the case at hand. It tells us how *BA* could be nontrivial for someone in Fred's position, it does not show that it is nontrivial for those in our position.

You might reply that the impression of *BA*'s nontriviality just is the impression that it would be nontrivial for someone in Fred's position Grant this, and the metalinguistic line solves the paradox, but at a price It dashes the hope that philosophers searching for analyses of truth, goodness, knowledge, and so on, are seeking more than merely linguistic enlightenment We would like to think that we are after *goodness*—for example—not just 'goodness' If the metalinguistic line is the only defensible solution to the paradox, analytic philosophers must be seeking mere linguistic fluency Such a verdict would deflect the charge of stupidity, which does not stick to mere inarticulateness The charge of shallowness, however, would still haunt the practice of analysis I will offer another solution that allows analysis greater depth Surprisingly, it involves embracing the triviality of *BA*

II EMBRACING TRIVIALITY, PART 1 TRIVIALITY AND UNDERSTANDING

I grant synonymy to 'boy' and 'male human child' (as used in some particular but presumably typical context) and I suppose that their common meaning is not metalinguistic (except where explicitly stated otherwise) Given this qualification, I allow that *BA* is trivial The paradox of analysis is sound—but innocuous

My defense of these claims has two main parts The first part (occupying the rest of this section) diagnoses the appearance that *BA* is nontrivial The second part (occupying the next section) aims at showing how analysis can be philosophically fruitful despite *PA*'s truth

BA is, I claim, "trivial " What do I mean? It is time to clarify this central notion

I distinguish three kinds of things that can be trivial propositions, sentences, and sentential schemas The kind of triviality at issue applies fundamentally to propositions and only derivatively to things of the other two kinds To say that a proposition is trivial (in the sense relevant to the paradox) is to say that it cannot be grasped without being truly believed[5] This characterization captures the idea that triviality is a matter of uninformativeness if you can grasp a true proposition without believing it, then you have something to learn If understanding that *P* suffices for truly believing that *P*, then you cannot wonder whether *P*, and this impossibility of wonder, coupled with truth, makes for triviality The proposition expressed by *BT* is an example of a *trivial proposition*

---

[5] My 'trivial' is short for 'trivially true', there is trivial falsity as well A weaker sense of 'trivial' makes understanding sufficient for (merely) believing I hope there are no trivial but false understandable propositions, on this weaker sense, but they do not worry us where the paradox of analysis is concerned

It is one thing to define triviality this way, it is another to assert that some propositions are trivial given this definition. I claim that the proposition expressed by *BT* is trivial. Thus I should answer Alvin Goldman's argument to the effect that no propositions are trivial in my sense. He says:

> Any proposition, we may assume, has two or more components that are somehow organized or juxtaposed. To understand the proposition one must 'grasp' the components and their juxtaposition. Now in the case of *complex* logical truths, there are (human) psychological operations that suffice to grasp the components and their juxtaposition but do not suffice to produce a belief that the proposition is true. But can't we at least *conceive* of an analogous set of psychological operations even for simple truths, operations which perhaps are not in the repertoire of human cognizers but which might be in the repertoire of some conceivable beings? That is, can't we conceive of psychological operations that would suffice to grasp the components and componential juxtaposition of these simple propositions but do not suffice to produce *belief* in the propositions? I think we can conceive of such operations. Hence for any proposition you choose, it will be possible for it to be understood without being believed.[6]

I offer two rebuttals—the second more generally acceptable, I suspect, than the first. First, I reject Goldman's assumption that "Any proposition . . . has two or more components that are somehow organized or juxtaposed." Propositions, as I understand them, occupy no space, and I can make no sense of a true, literal application of the notion of proper (nontemporal) components or parts to things that occupy no space. Others apparently can, and some construct elaborate theories of propositional structure. These theories simply assume, though, that propositions can have structure. Without a defense of this assumption, they stall on takeoff.

Perhaps I am blind to a possibility that other, cognitively better endowed philosophers see. From my unseeing perspective, though, it is more likely that the appearance of sight deceives. I suspect it stems from a conflation of figurative with literal meaning. Frege, whose principle of compositionality may seem to be the wellspring of the compound-proposition view, was well aware of this conflation:

> To be sure, we really talk figuratively when we transfer the relation of whole and part to thoughts [Frege's thoughts, our propositions], yet the analogy is so ready to hand and so generally valid that we are hardly ever bothered by the hitches which occur from time to time.[7]

---

[6] Alvin Goldman, "What Is Justified Belief?" in George Pappas, ed., *Justification and Knowledge* (Boston: Reidel, 1979), pp. 1–23; see p. 4.

[7] Gottlob Frege, "Compound Thoughts," translated by R.H. Stoothoff, *Mind*, LXXII (1963): 1–17. This article was originally published in German under the title "Logische

Propositions occupy no space, so they lack parts Had they parts, per impossibile, these could not be juxtaposed, since juxtaposition is an essentially spatial notion

Propositions have figurative parts—*being a boy* "is a part of" the proposition that all boys love broccoli The cash value of this claim, though, is just that it is impossible to believe that all boys love broccoli without grasping *being a boy* The part is proper insofar as the reverse does not hold grasping the concept does not require believing the proposition Linguistic representations of propositions have literal parts,[8] and claims about these may be taken (as Frege takes them) as figurative stand-ins for claims about propositional properties (Another reason these claims must be figurative if true distinct linguistic expressions with different parts may equally well represent a single proposition, and there is no basis for preferring one of these to others as revealing the proposition's "real" structure) From these and other cash equivalents of papery talk about propositional components, though, nothing like Goldman's argument gets off the ground

My second rebuttal, independent of the first and concessive of propositional structure, is that Goldman's argument is either invalid or question-begging Grant for argument's sake that to understand a proposition "one must 'grasp' [its] components and their juxtaposition " Grant, that is, that propositions have proper parts, and that grasping the parts of a proposition and their arrangement is necessary for grasping the proposition Grant, too, the assumption that even for very simple propositions we can "conceive of psychological operations that would suffice to grasp the components and componential juxtaposition of these simple propositions but do not suffice to produce *belief* in the propositions " Goldman infers fallaciously that any proposition can be understood without being believed What is needed to fix the inference is a premise to the effect that grasping the components and componential juxtaposition of a proposition suffices for understanding the proposition This sufficient condition for propositional understanding is distinct from the necessary condition that Goldman asserts

It is also dubious It is an old point that (even allowing structured propositions) a proposition cannot be identified with a mere concatenation of elements—say, an object and a property For such a concatenation does not tell us that the object instantiates the property (It

---

Untersuchungen Dritter Teil Gedankengefüge," in the *Beitrage zur Philosophie des deutschen Idealismus* (1923) 36–51

[8] Representation-tokens, that is

could just as well tell us that it does not, or that nothing else does, or any one of myriad other things) Goldman needs a sufficient condition where he offers only a necessary one His argument, taken at face value, is invalid

Perhaps we should understand the claim that we can "conceive of psychological operations that would suffice to grasp the components and componential juxtaposition of propositions but do not suffice to produce *belief* in the propositions," with less fidelity but greater charity, to mean simply that we can conceive of understanding the proposition without believing it In that case, Goldman's argument has the validity and vacuity of any argument whose conclusion merely repeats its premise For on this rendering Goldman is simply asserting that any proposition can be conceived to be understood without being believed, and any appearance of substantive argument is delusory

I offer no positive argument for my claim, which strikes me as obvious, that the proposition expressed by *BT* is trivial I am open to argument against the seemingly obvious Goldman's argument fails, though, and I know of no other argument that succeeds in its place

I return to triviality The triviality of a proposition is an absolute matter, but that of a sentence is relational To say that a sentence is trivial is to say that the proposition it expresses is trivial Roughly, anyway a sentence can express different propositions to different people—and to the same person at different times—some trivial and some not Sentential triviality must be relativized, strictly speaking, to particular uses of sentence-tokens Given my background assumptions about the context and uses of (tokens of) *BA* and *BT* under consideration, however, I will relax and say that *BT*, by virtue of expressing a trivial proposition in all its standard uses, is a *trivial sentence* (With 'standard uses' I exclude those in which *BT* is used as part of a code, for example, and those in which its two occurrences of 'boy' are nonsynonymous) More generally, we can say that a sentence (-type) is trivial just in case all standard uses of its tokens express trivial propositions

You might challenge the propositionalism on which my characterization of triviality depends Those who balk at propositions typically indulge those of us who buy them, boasting that everything we can say, they can say better If this boast can be made good, then merely rejecting propositions undermines neither the paradox of analysis nor my diagnostic response to follow What is needed instead is a propositionless reconstrual of synonymy that first explains the initial plausibility of (1)–(6) and then undercuts the plausibility of at least one of them I know of no such reconstrual

You might instead challenge the claim that triviality applies ultimately to propositions, without denying propositionalism Perhaps to say that a sentence is trivial is to interpose something between the believer and the proposition expressed by the sentence—the sentence itself, perhaps, or a nonsentential "aspect" under which a proposition is believed—such that the triviality in question attaches ultimately rather than derivatively to this intermediary Two questions then arise Is the alleged intermediary intelligible, and if so, what is it for the intermediary to be trivial, where this is not explicable in terms of the triviality of the proposition?

Ernest Sosa suggests a solution along these lines [9] Sosa would rightly embrace the idea that to believe the proposition expressed by *BT* is to believe that expressed by *BA* Unwilling to accept the triviality of *BA*, he invokes something else for the sole purpose of making two beliefs out of one Given the *ad hoc* nature of the something else, I for one have no real grasp of what it is supposed to be [10]

Jeffrey King gives an account of analysis appealing to structured abstract propositions [11] As I have said, I do not see how nonspatial propositions can have literal structure, and I do not see how a nonliteral interpretation of King's suggestion might solve the paradox of analysis The acceptance of views like King's merits more discussion than I can offer here, but it should be clear that they face at least an apparent problem of intelligibility

Even more importantly, though, my solution to the paradox does not hinge on whether views like Sosa's and King's can be made intelligible and defensible There are two reasons First, I offer a solution to the paradox that renders such views superfluous, since it invokes no suspect entities or attributes that they do not while omitting some that they include, hence it beats them for ontological thrift Second, even if such views identify a flaw in the paradox, mine identifies an independent one, and mine, unlike theirs, also provides a crucially needed diagnostic explanation for the misleading appeal of the mistaken claim

I return again to triviality and consider a third kind, that of a *trivial sentential schema* A sentential schema is trivial (in my stipulated sense) if and only if any uniform, grammatical substitution of meaningful terms[12] for its variables yields a trivial sentence [13]

---

[9] Sosa, "Classical Analysis," this JOURNAL, LXXX, 11 (November 1983) 695–710

[10] I am not alone, see George Bealer, "Remarks on Classical Analysis," this JOURNAL, LXXX, 11 (November 1983) 711–12

[11] King, "What Is a Philosophical Analysis?" *Philosophical Studies*, XC, 2 (1998) 155–79

[12] "Uniform substitution" requires uniformity of meaning, not merely of syntactic-type

[13] To say that a term is "meaningful" may be to say that it has a sense, a reference, or merely a character For present purposes, only sense and reference count

'A *G* is a *G*' is a sentential schema, not a sentence It is a trivial schema substitute meaningfully and uniformly for '*G*', preserving grammaticality, and you have a trivial sentence Since the triviality of sentential schemas derives from that of sentences, and that of sentences from that of propositions, the triviality of sentential schemas derives from that of propositions

Propositions, sentences, and schemas may properly be considered trivial, though triviality is different for each Consider now another notion that may apply to sentences, *schematic triviality* (Warning! Schematic triviality is not triviality, and it applies to sentences, not to schemas) To say that a sentence is schematically trivial is to say that it is a substitution-instance of a trivial sentential schema—but *merely* that it is a substitution-instance, not that it results from meaningful uniform substitution

Consider the following sentences

(a) A froe is a froe
(b) No furkels are nonfurkels
(c) If snaths exist, then some snaths are snaths
(d) If he did it, then he did it
(e) If the first word of this sentence is 'the', then the first word of this sentence is 'the'

Which of these sentences are trivial? If you say that all of them are, you are naturally and interestingly mistaken All are schematically trivial, but only (a) and (c) are trivial, and if the meanings of 'froe' and 'snath' escape you, then you know this only by taking my word for it

Any competent English speaker will recognize that sentences (a)–(e) are trivial—that they express trivial propositions—*if they are uniformly meaningful*, if, for example, the two occurrences of 'froe' in (a) both mean something and both mean the same thing And they express trivial propositions, if they are uniformly meaningful, because they are uniformly meaningful substitution-instances of trivial sentential schemas In other words, (a)–(e) are schematically trivial

The qualification 'if they are uniformly meaningful' is crucial 'Furkel' is a nonsense word, and (b) expresses no proposition whatsoever So it does not express a trivial proposition and thus is not a trivial sentence It is nevertheless schematically trivial What determines whether a sentential schema is trivial is not whether its substitution instances are trivial, but whether they are trivial when they result from uniform, grammatical substitution of meaningful terms for schematic variables Thus a trivial schema can have nontrivial substitution-instances—those that express no proposition, for example

These are not, however, the only cases of trivial schemas with nontrivial substitution-instances Consider (e) If its two occurrences of 'this sentence' refer to the same sentence—whether it is the antecedent of (e), the consequent, or (e) itself—then (e) expresses a trivial proposition If the first occurrence refers to the antecedent but the second to (e) itself, however, then (e) expresses a false—hence nontrivial—proposition Since any of these uses could equally be considered "standard"—whether all could be or none could be—(e) itself is not trivial It is not nonsense, though And it is a substitution-instance—though not on all standard uses a uniformly meaningful substitution-instance—of a trivial sentential schema 'If *P* then *P*'

You might object—especially if you like your languages formal—that (b) is not an English sentence at all But on what grounds can you exclude it? English, a natural rather than a formal language, has no fixed lexicon, it permits coinages, including nonsense coinages [14] Competent English speakers are adept at distinguishing English nonsense sentences from nonsentential nonsense like 'dipohsg shgu hsdui' At least this is the most natural description of the phenomenon, and there is no good reason to recant it [15] I need not press the point here, though, since (d) and (e) will do for my purposes The nonsense sentence helps bring home the distinction between a trivial schema and schematic triviality That distinction does not depend, though, on the existence of nonsensical but genuine sentences

Triviality and schematic triviality are independent a sentence may be schematically trivial without being trivial, and vice versa Furthermore, a sentence may be trivial even though competent speakers do not know that it is, as at least one of (a) and (c) is likely to illustrate [16] Finally, schematic triviality, unlike triviality, is not closed under synonymous substitution, as *BA* and *BT* show Not every sentence synonymous with a schematically trivial sentence is schematically trivial

These distinctions help to explain the initial judgment that *BA* is not trivial while *BT* is, and they help to dispel the air of paradox surrounding the claim that *BA is* trivial. What is clearly true is that *BA* is not, while *BT* is, schematically trivial. Schematic triviality is easily confused with triviality and not easily distinguished from it (even by some respected philosophers who will go unnamed here). Anyone prone to this natural confusion will incline to the idea that *BA* is nontrivial.

---

[14] Some standard English words are nonsense—consider 'widget' Other nonsense coinages have gained meaning 'chortle' and 'whiffle', for example, first used in Lewis Carroll's *Jabberwocky*

[15] Here is a bad reason (from correspondence) "You say, in effect, that replacing a meaningful word in a sentence with a nonsense word leaves a sentence Keep applying this principle, and you'll end up with a 'sentence' comprising *only* nonsense words But this is absurd " Reply I have not endorsed the soritical principle that any substitution of nonsense for sense in a sentence preserves sentencehood I simply claim, on the basis of clear examples, that it sometimes does

[16] Furthermore, you can know that a sentence is trivial without recognizing it as trivial You may know that (a) expresses a trivial proposition—say, on my testimony—without grasping that proposition

If your judgments about the applicability of 'trivial' track schematic triviality rather than triviality, you will embrace both (1) and (6). And you will find a "solution" to the paradox in rejecting (5)—rightly, in a way, since (5) is false if we read its 'triviality' as meaning *schematic triviality*.

If 'triviality' meant *schematic triviality*, then, we could solve the paradox without granting *PA*. But 'triviality' does not mean this. Sentential triviality, unlike schematic, is closed under the synonymy that conceptual analysis yields. And if the meaning of the terms in question is not metalinguistic, the paradox of analysis is sound.

III. EMBRACING TRIVIALITY, PART 2: THE ACTIVITY AND RESULT OF ANALYSIS

I have diagnosed the misleading pretheoretical appeal of (6); the paradox is solved. What remains is to offer a salve to the philosophical spirit by showing that accepting *PA* does not condemn analyzing philosophers as shallow and stupid.

'Analysis' is ambiguous. Sometimes it refers to a sentence that purportedly provides a meaning-analysis, such as *BA*. Call this an *analyzing sentence*. The "conceptual analysis" mentioned in *PA* is an analyzing sentence. Sometimes 'analysis' refers to *analyzing*, the philosophical activity that aims to, and sometimes does, culminate in the production of a true analyzing sentence.

I need not offer an account of what it is to be an analyzing sentence to make my case, but a rough account will be helpful, so here it is. An analyzing sentence is (roughly) one that explicitly represents the analysandum as a conjunction of conditions such that (a) the conditions are individually necessary and jointly sufficient for the analysandum but not individually sufficient, (b) the conditions are individually graspable without any grasp of the analysandum, and (c) grasping the conjunction of the conditions is grasping the analysandum. *BA* illustrates. It represents *being a boy* as a conjunction of three conditions—*being male*, *being human*, and *being a child*,[17] these conditions are individually insufficient, individually necessary, and jointly sufficient for being a boy, and grasping the conjunction of these conditions—grasping *being male and human and a child*—just is grasping *being a boy*

---

[17] It may look here as though I identify conditions with concepts. I do not, in fact, mind doing this for present purposes, though other interpretations of 'condition' are available. This is one reason why this account of analyzing sentences is rough. For one step in the direction of polish, I will say that given the standard interpretation of the predicate calculus, it is linguistically sufficient for *S*'s being an analyzing sentence that

My condition (c)—that grasping the conjunction of the *analysans*-conditions just is grasping the *analysandum*—is needed to distinguish so-called "criteriological" analysis from conceptual analysis, only the latter poses an apparent paradox worthy of the name [18] This condition also effectively guarantees the truth of *PA* [19]

Notice that I say that an analyzing sentence represents its *analysandum* conjunctively, observing my earlier strictures, I do not say that the *analysandum is* a conjunction Thus "a conceptual analysis" is not a proposition but a sentence Two sentences may be synonymous while only one of them counts as an analyzing sentence—a conceptual analysis—by virtue of explicitly representing the *analysandum* conjunctively See, for example, *BA* and *BT*

You know what it is for an analyzing sentence to be trivial To say that analyzing is trivial is to say that it yields no new insight, knowledge, or understanding, otherwise it is nontrivial or enlightening Given that all true analyzing sentences are trivial, it hardly follows that analyzing is trivial

Suppose you analyze and succeed Your result is a trivial analyzing sentence If you grasped the analyzandum-concept before you began, you have gained no grasp of that concept But you may not have grasped it, you may not acquire the concept until you arrive at the trivial analyzing sentence

---

(i) *S* is a substitution instance of '$(x)(Fx$ iff $(Gx$ & $Hx))$',

(ii) the substituends for '*F*', '*G*', and '*H*' are each univocal and express distinct concepts,

(iii) the concept expressed by the substituend for '*G*' can be grasped by someone who does not grasp that expressed by the substituend for '*H*', and vice versa, and

(iv) all substitution instances of '*Fa*' and of '*Ga* & *Ha*' are synonymous

These sufficient conditions are not necessary, and they retain roughness, but they should convey the notion of an analyzing sentence

[18] Felicia Ackerman responds to the paradox of analysis by arguing, in effect, that a true analysis cannot be conceptual but must instead be criteriological See "The Informativeness of Philosophical Analysis," in Peter French et al , eds *Midwest Studies in Philosophy*, Volume VI (Minneapolis Minnesota UP, 1981), pp 313–20, and "Essential Properties and Philosophical Analysis," in French et al , eds , *Midwest Studies in Philosophy*, Volume XI (Minneapolis Minnesota UP, 1986), pp 304–13

[19] Condition (c) means, as applied to the case of *BA*, that the concept of grasping the concept of being a boy is numerically identical with the concept of grasping the concept of being human and male and a child

You might acknowledge this as an abstract possibility but doubt its practical applicability For when philosophers try to analyze (say) knowledge or freedom, there is a presumption that we grasp the concept from the outset—else how could we recognize either successful or unsuccessful analysis?

In fact, you need not possess a concept in advance of producing a successful analyzing sentence You need merely possess something that tracks it A persistent theme in contemporary philosophy of language is that the reference of some uses of referring terms is guided by social facts about other uses of those very terms For example, the reference of Hilary Putnam's use of 'beech' is famously fixed in part by the way 'beech' is used by those in the arboreal know, experts whose own use is not determined by such metalinguistic facts [20] No one disputes this nowadays, though conflict continues over whether the metalinguistic conditions that guide reference are part of the sense Putnam associates with 'beech'

How does this bear on analysis? Someone's use of a term might be guided by facts about her own disposition to apply it Suzy might apply 'knowledge' to a case simply because she finds herself disposed to call it 'knowledge' This disposition need not rest on a grasp of any nonmetalinguistic concept that 'knowledge' expresses, though the disposition may nevertheless track such a concept The tracking might occur either because she has a nonconceptual responsiveness to knowledge or because she associates a partially metalinguistic concept with 'knowledge' that tracks the nonmetalinguistic concept of knowledge These two options mirror the semantic externalist and internalist options for Putnam's use of 'beech' On either option, if Suzy seeks and finds a true analyzing sentence, two things may happen simultaneously she acquires the (nonmetalinguistic) concept of knowledge and grasps the trivial proposition expressed by the analyzing sentence She is enlightened [21]

---

[20] Putnam, "The Meaning of 'Meaning'," in *Mind, Language and Reality Philosophical Papers, Volume 2* (New York Cambridge, 1975), pp 215–71, reprinted from K Gunderson, ed , *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science,* Volume VII (Minneapolis Minnesota UP, 1975) Putnam says that his "*concept* of an elm tree is exactly the same as [his] concept of a beech" and adds that this shows it a mistake to identify meaning "in the sense of intension" with concept (p 226) It shows this a mistake only if "concepts" are construed as mental entities, rather than as the abstract, intensional entities that pose the paradox of analysis, as I take them to be

[21] "Objection Let *KT* = 'Knowledge is knowledge', and let *KA* be a true analyzing sentence for knowledge If we acquire the nonmetalinguistic concept of knowledge from *KA*, then we acquire the same concept from *KT* This is no 'major conceptual advance' Since we already have the constituent concepts by means of *KT,* any advance

Analyzing may in some cases be akin to ostension. Consider how a child acquires the concept *chair*. It is not by analytic definition. The child sees chairs, hears them called 'chairs', and calls things 'chairs' but without grasping the concept of chairhood. Then, somehow, she gets it. The etiological and psychological details of this understanding remain mostly mysterious, but the fact of it is plain. Analyzing may be an etiology of understanding that is relevantly like the etiology of understanding via ostension.

Am I then claiming that we do *not* grasp nonmetalinguistic concepts expressed by 'knowledge' or 'freedom' before successful analysis? If so, then I am stuck with denying that we grasp such concepts, since I deny that we possess any true analyzing sentences for these concepts. But I am not stuck with this denial. All I have claimed is that if we do not believe the proposition expressed by a true analyzing sentence, then we do not possess (at least one of) the concepts expressed by a constituent of that sentence. But if there is no true analyzing sentence to be had, we may perfectly well possess all the relevant concepts. I may perfectly well possess the concept of knowledge despite failing to believe the proposition expressed by a true analyzing sentence of the form '*S* knows that *p* if and only if ...', for there may be no such sentence. The concept of knowledge may be unanalyzable.

For any concept, we can easily give conditions that are individually necessary, individually insufficient, but jointly sufficient for its application.[22] It may even be possible for any concept to give an informative criteriological analysis. It is hardly obvious or plausible, though, that all concepts permit conceptual analysis. If knowledge is unanalyzable, then there is no correct analyzing sentence of the form '*S* knows that *P* if and only if ...', assuming that the use of 'knows' in such a sentence would mean *knows*. Our failure to bag an analyzing sentence may be explained by the non-existence of our quarry.

I do not say that there is no proposition that such a sentence would express if it did exist. There is such a proposition: it is the proposition

---

is merely linguistic (or metalinguistic), for it involves no more than putting words to those concepts." Response: I deny that we must have the relevant concepts by means of *KT*. If there is a true analyzing sentence for knowledge, and I do not now understand the proposition it expresses, then I do not now understand *KT* despite its schematic triviality (and I do not now understand this very sentence), for I do not now understand 'knowledge'.

[22] Here is one helpful schema: '*a* is *F* iff (1) *a* is *F* or *G*, (2) *a* is *F* or *H*, and (3) *a* is *F* or (not-*G* and not-*H*)'. Note, however, that substitution-instances of this schema will be analyzing sentences only if they satisfy the "grasping"-conditions on an analyzing sentence—no mean feat. Merely substituting for '*F*' on the right side of the biconditional a synonym for the substituend on the left side will not do the trick.

that *S* knows that *P* if and only if *S* knows that *P*[23] I believe this proposition But (to repeat) I nevertheless fail to believe the proposition *actually expressed by a true analyzing sentence for knowledge*, if there is no such sentence, even though I believe the proposition that such a sentence would express if it existed

Am I committed to the claim that if tomorrow we discover a true analyzing sentence for knowledge, then yesterday we did not grasp a (nonmetalinguistic) concept of knowledge? No, since the possibility remains that what is gained is not understanding of knowledge but of the words used in the analyzing sentence, the metalinguistic line *may* be right Fred may know perfectly well that a boy is a male human child but not know until he confronts *BA* that (say) 'human' means human Since I have no analyzing sentence for knowledge, I cannot give you the analogous case, but the idea is clear

If what is gained is not merely metalinguistic knowledge—and I have shown how such a gain is possible—then yes successful analyzing would yield the possession of a new concept We would acquire the nonmetalinguistic concept of knowledge, a concept whose instantiation would be tracked by the linguistic disposition or partly metalinguistic concept that we associated with 'knowledge' yesterday And this would be a major conceptual advance, one that vindicates the philosophical value of analyzing Of course, all of this rests on a very big "if"

So much for successful analyzing What about unsuccessful? Suppose you try and fail to find a correct analyzing sentence for 'knowledge' Suppose, too, that you know you have failed Then you should have learned a disjunctive truth (a) there is no analyzing sentence for the concept you associate with 'knowledge', or (b) your use of 'knowledge' is guided partly by a nonconceptual responsiveness to knowledge, or (c) the concept you associate with 'knowledge' (if any) is partly metalinguistic, or (d) there are concepts you grasp for which you lack words

This disjunction follows from the two available explanations for your failure One is that there is no true analyzing sentence to be had In this case, (a) is true I have already discussed this possibility The other explanation is that there is a true analyzing sentence to be had but we have not yet found it In this case, at least one of (b), (c), and (d) is true

How can we choose among the disjuncts? This is a hard epistemic question, and inability to answer it does not signal stupidity Answering it exceeds my ambition (and probably my ability) here Raising it, however, helps satisfy that ambition For in seeing how this question arises from the failure of analysis, you can see how even failed analyzing can raise deep questions about our understanding If we are left wondering what explains our failure, then we are left wondering which of (a)–(d) is true We may seek resolution of this wonderment A discovery that any of (a)–(c) is true would have direct philosophical value, and even discovering (d) to be true might have indirect value The realization that a philosophically valuable discovery may await us in connection with failed analysis is itself a philosophically valuable realization Persistent failure to satisfy our wonderment, if such we confront, will reveal some of our cognitive limitations Such revelation, too, has at least indirect philosophical value

---

[23] '*S* knows that *P* if and only if *S* knows that *P*' expresses no proposition, but a suitably quantified version of it—where '*S*' ranges over people and '*P*' over univocally meaningful sentences—does

Other puzzles of analysis await. What should we say, for example, about those who embrace a mistaken analyzing sentence? We should say that they are confused, of course, but the nature of such confusion is itself a matter of philosophical perplexity I will not address this issue or related ones here, for they are tangential to my chief aim of vindicating analysis while solving its paradox

I confess dubiety about analyzing philosophically interesting concepts I suspect that such concepts—like most concepts—are unanalyzable This confession of attitude is no argument I mention it only to emphasize that even we who have this attitude and who generally eschew conceptual analysis can consistently recognize its value For unless this attitude of dubiety hardens into positive conviction—a transformation for which I see no justification—we must recognize the real chance that our doubts are mistaken To recognize this chance is to recognize that the activity of analysis, whether successful or not, has the potential for deep and hard won philosophical insight Some analytical philosophers are shallow and stupid Their failings are individual ones, however, and not occupational inevitabilities

EUGENE MILLS

Virginia Commonwealth University

GUILT BY STATISTICAL ASSOCIATION REVISITING THE PROSECUTOR'S FALLACY AND THE INTERROGATOR'S FALLACY

The "prosecutor's fallacy" and the "interrogator's fallacy" are names for two different and very specific ways of inferring a suspect's guilt Although I do agree that these two kinds of reasoning involve some elementary logical mistakes, I think that in exposing these fallacies the critics sometimes introduce serious misconceptions of their own They try to educate the wider public about how to deal with these probability issues but in the end they just manage to muddy the waters even more with their logically flawed analysis In particular, I will try to show that in each of these two situations the critics fail to recognize some legitimate probabilistic indicators of guilt that remain available for a nonfallacious inference My diagnosis of the problems in the two cases will be structurally very similar, and this explains why they are being discussed together in this article

I THE PROSECUTOR'S FALLACY

The prosecutor's fallacy is a conflation of two conditional probabilities Let M be the statement that a defendant's DNA matches the DNA of the perpetrator (obtained from the crime scene), and let I be the statement that the defendant is innocent If $p(M/I)$ is very low (that is, it is very improbable that there would be a DNA match if the defendant is innocent), this is sometimes taken to mean that $p(I/M)$ is also very low (that is, that it is very improbable that the defendant is innocent if there is a DNA match) This is an obvious mistake because it is clear that the two probabilities are logically quite distinct and that neither can be directly inferred from the other

The confusion of $p(M/I)$ and $p(I/M)$ is called the "prosecutor's fallacy" because the prosecutor is often interested in establishing that $p(I/M)$ is very low (which would indeed amount to strong evidence that the defendant is guilty), but occasionally what he instead presents to the court as incriminating evidence is just the fact that $p(M/I)$ is very low, which *in itself* does not point to the guilt of the defendant at all

It is easy to expose this error but unfortunately it is also easy to commit another error in the process Sometimes the criticism of the prosecutor's fallacy is taken too far and the evidential relevance of the DNA match for the defendant's guilt is not recognized even when it does have probative value In some cases, surprisingly, even distinguished statisticians fall into this trap

0022-362X/08/0506/320-32

Let us start with Gerd Gigerenzer, one of the currently most prominent experts in the application of probability to practical problems. In explaining the perils of the prosecutor's fallacy, he discusses a criminal case from Germany in which two pieces of evidence were adduced against a defendant.[1] First, the defendant's blood matched the blood found under the fingernails of a murdered woman, with this particular blood type being shared by 17.3 percent of Germans. Second, a different blood type was found on the defendant's boots and it matched the blood of the woman, with this blood type being shared by 15.7 percent of Germans. Now how probable is it that the defendant is guilty if only this evidence is taken into account? Gigerenzer addresses the question and offers an answer that is inadequate on multiple levels.

He starts with estimating $p(M/I)$, that is, the probability of the blood match, given that the defendant is innocent. Gigerenzer suggests that we calculate the probability of the double match (first, between the suspect's blood and the blood under the woman's fingernails, and second, between the blood on the suspect's boots and the woman's blood) by simply multiplying the probabilities of each match (*ibid.*, p. 157). But this is a patently wrong approach to the second match. The fact that 15.7 percent of people share the same blood with the victim surely cannot mean that the probability of an innocent person having this blood type *on his boots* is also 15.7 percent! It must be much, much lower, simply because it is extremely rare that an innocent person has someone else's blood on his boots (or anywhere on his clothes, for that matter).

After allowing this ill-considered multiplication of probabilities (17.3 percent × 15.7 percent) Gigerenzer concludes that the probability of the double match (given innocence) is 2.7 percent. So $p(M/I) = 0.027$. But what is the numerical value of $p(I/M)$? Those who commit the prosecutor's fallacy would say that it is the same, and from this it would follow that the probability of guilt, $p(\sim I/M)$, is overwhelming, namely, higher than 0.97.

This is of course wrong but even if we forget the problems with Gigerenzer's unhappy multiplication of probabilities his estimate of $p(\sim I/M)$ is still not correct. He first assumes that any of roughly 100,000 men in the city could have committed the crime. Of all these men, 2.7 percent will have the double blood match, which means around 2,700 of them. Now since the defendant belongs to a group of 2,700 men, and one member of that group is the perpetrator, the probability that the defendant is the murderer is 1 in 2,700. Therefore, according to him, $p(\sim I/M)$ is not 0.97 but 0.0004.

---

[1] Gigerenzer, *Reckoning with Risk: Learning to Live with Uncertainty* (New York: Penguin, 2002), pp. 156–58.

Interestingly, in another text[2] in which he discusses the same case, Gigerenzer is more careful and he realizes that the probability of someone's having the victim's blood on his boots cannot be equal to the proportion of people with that particular blood type in general population. He is aware that this probability must be much lower (because not everyone has someone else's blood on him) but, oddly enough, he now widens the range of suspects from the previous 100,000 to 10 million, and by fudging the numbers in this way he makes sure that the probability of the suspect's guilt given the DNA match again comes out very low.

Before I point out the central mistake in Gigerenzer's reasoning let me present a similar analysis of another case (*R. v. Adams*) that is offered by the Oxford statistician Peter Donnelly.[3] A man was convicted of raping a woman, and again the only incriminating evidence against him was the fact that his DNA profile matched the DNA evidence from the crime scene. Donnelly claims that even if this DNA type were so rare that it is found only in one person in 20 million, this would still not make it more probable that the defendant was guilty than that he was not.

His argument is essentially the same as Gigerenzer's. From the assumption that $p(M/I)$ is 1 in 20 million Donnelly concludes: "If you believe that number, then on average there will be 2 or 3 people in Britain whose DNA it could be, and probably no more than 6 or 7" (*ibid.*, p. 48). It would seem then that despite the extremely low probability of the DNA match (on the assumption of innocence), the probability of the defendant's guilt is still not very high even after it is established that his DNA matches the DNA of the perpetrator.

Although Donnelly does not provide a mathematical derivation of that result, there is a simple Bayesian formula that underlies his argument. The goal is to obtain the probability of guilt, or $p(G/M)$, given the DNA match. The calculation is straightforward because all four necessary elements are known:

$p(G)$ = the probability of an arbitrarily chosen person being guilty = 1 in 60 million (all people in Britain)

$p(\sim G)$ = the probability of an arbitrarily chosen person not being guilty = $1 - p(G)$, that is, very close to 1

---

[2] Gigerenzer, "Ecological Intelligence: An Adaptation for Frequencies," in D. D. Cummins and C. Allen, eds., *The Evolution of Mind* (New York: Oxford, 1998), pp. 9–29.

[3] Donnelly, "Appealing Statistics," *Significance*, ii (2005): 46–48.

$p(M/G)$ = the probability of the DNA match if the person is guilty = 1

$p(M/{\sim}G)$ = the probability of the DNA match if the person is not guilty = 1 in 20 million

$$(1)\qquad p(G/M) = \frac{p(G) \times p(M/G)}{p(G) \times p(M/G) + p({\sim}G) \times p(M/{\sim}G)}$$

$$= \frac{\frac{1}{60m} \times 1}{\frac{1}{60m} \times 1 + \approx 1 \times \frac{1}{20m}} = 0.25$$

Donnelly is aware, of course, that he neglects some relevant considerations, and that many people in Britain would obviously be excluded from suspicion on a number of grounds (like gender or age). Yet although he explicitly warns that the class of potential suspects is for this reason much narrower than the whole population of the United Kingdom and although it is quite easy to introduce this correction, Donnelly inexplicably proceeds to work out his calculation on the basis of this very unrealistic assumption. The result (the probability of 0.25, obtained in (1)), is a serious underestimation of the true probability of guilt. For if we first exclude all women (30 million) and then also a number of men that are ruled out on grounds like age and so on (probably around 20 million), we are left with the group of potential suspects of approximately 10 million. As equation (2) shows, doing the calculation on the basis of this much more plausible assumption, p(G/M) turns out to be considerably higher than previously, namely 0.66.

$$(2)\qquad p(G/M) = \frac{p(G) \times p(M/G)}{p(G) \times p(M/G) + p({\sim}G) \times p(M/{\sim}G)}$$

$$= \frac{\frac{1}{10m} \times 1}{\frac{1}{10m} \times 1 + \approx 1 \times \frac{1}{20m}} = 0.66$$

Yet there is a more serious mistake lurking in Donnelly's and Gigerenzer's approach to the prosecutor's fallacy. Although they are right that the simple transposition of two probabilities, *p*(*M*/*I*) and *p*(*I*/*M*), is a fallacy, I will argue that in each of their two cases the DNA match actually does constitute much stronger evidence that the defendant is guilty than they recognize. The basic mistake in their analysis consists in disregarding a crucial aspect of the situation, which should have been taken into account and which indeed significantly lowers the estimated value of *p*(*I*/*M*), and consequently increases the probability of the defendant's guilt.

The fact that in *R v Adams* the defendant's DNA was found to match the DNA from the crime scene entails that his DNA *was known to the police* That is, it was available in the database that the police used in looking for a match And precisely this piece of information is the dog that barked loudly but was ignored by both Gigerenzer and Donnelly If it were true that the police databases contain DNA specimens of different people based on completely random sampling, the reasoning of the two statisticians would be perfectly fine But clearly this is not how things are It is common knowledge that the DNA included in police files predominantly belongs to individuals with previous convictions or arrests And, most importantly, it is well established that the probability of these people committing a crime is significantly higher than for the general population For instance, in the research on this matter that was recently conducted for the British Home Office, it is stated as completely uncontroversial that "generally the best predictor of future offending is a previous history of offending"[4]

So the situation is more complicated than presented in (1) and (2) For estimating the probability of the defendant's guilt *all things considered* there are *two* facts that have to be taken into consideration, not just one Besides M (the fact of the DNA match) there is also D (the fact that the defendant's DNA was in the police database) In other words, we are not looking for $p(G/M)$ but actually for $p(G/M\&D)$ Here is an appropriate equation for that purpose

(3)

$$p(G/M\,\&\,D) = \frac{p(G/D) \times p(M/G\,\&\,D)}{p(G/D) \times p(M/G\,\&\,D) + p(\sim G/D) \times p(M/\sim G\,\&\,D)}$$

There are four probabilities on the right side of the equation Using Donnelly's example again, the magnitudes of two of these probabilities are already known

$p(M/G\&D)$ = the probability of DNA match given that one is guilty and in database = 1

$p(M/{\sim}G\&D)$ = the probability of DNA match given that one is innocent and in database = 1 in 20 million

The probability of M given G and D is 1 because it was assumed from the beginning that the DNA collected at the crime scene did belong to the perpetrator And the probability of M given $\sim G$ and D is 1 in 20 million because if one is not guilty then the probability of one's DNA matching the perpetrator's DNA must be the same, *whether one's DNA is in a police database or not*.

---

[4] B Powis, *Offenders' Risk of Serious Harm A Literature Review* (London, UK Home Office, 2002), p v

How about the two remaining probabilities $p(G/D)$ and $p(\sim G/D)$? They have to add up to one, so we have to assign the value only to one of them. For the sake of the discussion, let us suppose first that the probability that a person with previous conviction(s) has 100 times higher probability of committing a crime than a randomly chosen person from the general population.[5] Then since the probability of an arbitrarily picked out individual committing this particular crime was assumed to be 1 in 10 million, we obtain the following values for the remaining two unknowns:

$p(G/D)$ = the probability of guilt given that one's DNA is in police database = 1 in 100,000

$p(\sim G/D)$ = the probability of innocence given that one's DNA is in database = 99,999 in 100,000

Substituting these values into equation (3) yields 0.99 as the value of $p(G/M\&D)$.

$$(4) \qquad p(G/M\,\&\,D) = \frac{\frac{1}{100{,}000} \times 1}{\frac{1}{100{,}000} \times 1 + \frac{99{,}999}{100{,}000} \times \frac{1}{20{,}000{,}000}} = 0.99$$

Alternatively, if we suppose that $p(G/D)$ is only fifty times (rather than a hundred times) higher than $p(G)$, equation (4) would still give the same answer for $p(G/M\&D)$ as before: 0.99. Even if $p(G/D)$ is set to be as low as just ten times higher than $p(G)$, $p(G/M\&D)$ would continue to be higher than 0.95.

In Gigerenzer's case as well, an analogical correction produces a similar surge in the key probability. Recall that in his example the value calculated for $p(G/M)$ was 0.0004. After necessary amendments are made in order to correct for the ludicrous multiplication of probabilities and after it is taken into account (in a similar way as with Donnelly's example) that those whose data are in police databases are significantly more likely to commit a crime, again the estimate of $p(G/M\&D)$ jumps from 0.0004 to values higher than 0.9.

---

[5] If some readers feel that this estimate might be a wild and unrealistic exaggeration let me point out that, for example, according to criminal statistics from England and Wales, a person convicted of a sexual offense (the crime that is relevant for our context) is between 80 and 250 times more likely to be convicted of a similar act in the first two years after being released than some randomly chosen individual. The source: M. Redmayne, "The Relevance of Bad Character," *Cambridge Law Journal*, LXI (2002): 695.

The fallacious reasoning that is criticized here has even crept into the pages of *Nature*. Statisticians David Balding and Donnelly consider a hypothetical case in which $p(M/I)$ is 0.000001, and they try to show that under certain assumptions $p(I/M)$ would still be relatively high, or 0.33 to be exact. Here is their reasoning: "There may be many individuals who, *if not for the DNA evidence, would be as likely to be the culprit as the defendant*. If there were 500,000 such individuals, for example in a large city, the probability that the defendant is innocent after taking the DNA evidence into account would be at least one-third."[6]

The italicized phrase is wrong because it implies that in the absence of the DNA match the defendant would be as likely to be the culprit as any member of the general population that possesses the general characteristics ascribed to the culprit (for example, male, age between 20 and 40, and so on). But in fact, as explained above, even without the DNA match the defendant's probability of guilt is already significantly higher than for the average person with those given characteristics, due to the presence of his DNA in police databases.[7]

Someone might object here that the increased probability of committing a crime that is inferred from the presence of one's DNA in police databases (and ultimately from the fact that these available DNA specimens massively belong to previous offenders) would be inadmissible evidence in a court of law and that Gigerenzer and Donnelly therefore justifiably ignored this information. There are three reasons why this defense is not convincing.

First, the previous offender status is not universally inadmissible for determining guilt. For instance, in some U.S. states (for example, Wisconsin) the prosecutor is allowed to introduce this kind of evidence in connection with some kinds of sexual offenses. And remember that both cases discussed by the statisticians involve rape. Moreover, some scholars argue that the criminal justice is already rapidly moving in the direction of allowing the evidence of bad character in criminal trials: "The winds of change look to be blowing towards radical reform of the policy which excludes evidence of a defendant's bad character. In the near future, evidence of previous misconduct may be presented to juries much more often than at present."[8]

---

[6] Balding and Donnelly, "How Convincing Is DNA Evidence?" *Nature*, CCCLXVIII (1994): 285 (italics added). Cf. Balding and Donnelly, "The Prosecutor's Fallacy and DNA Evidence," *Criminal Law Review* (1994): 719.

[7] Indeed, in the aforementioned rape case (*R. v Adams*), in which Donnelly was a witness for the defense, the defendant's DNA also became available to the police only after he was convicted of another, separate sexual assault. See "Rapist Jailed on Sole Evidence of DNA Database," *The Times* (September 14, 1996).

[8] Redmayne, "The Relevance of Bad Character," *Cambridge Law Journal*, LXI (2002): 714.

Second, the fact that this kind of evidence is formally inadmissible to the court does not mean that the jurors will not use it tacitly in their deliberations After all, in reaching a decision they are asked to use their common sense, and it is part of common-sense knowledge that DNA databases largely contain samples of previous offenders with increased crime risk Therefore, any statistician educating the public about the prosecutor's fallacy should address the impact of that piece of empirical knowledge on the estimation of relevant probabilities (even if he decides in the end to advise the reader to suppress this information in the *legal* determination of guilt) If this issue is completely ignored, the probability analysis offered by the statisticians will be unconvincing since it really omits a crucial factor that is highly relevant for the correct calculation of probabilities In other words, it may well be that the statisticians' solution encounters resistance and looks so counterintuitive and unacceptable not because people are confused about probabilities but rather because this "solution" proposed by the statisticians is just *wrong*, all things considered

Third, someone's guilt is often an issue in an out-of-court context The correct handling of probabilities is very important there as well, and it is not restricted by usual legal limitations Suppose that you learn that your children's baby-sitter has been interrogated by the police in a child-molesting case because his DNA matched the DNA collected at the crime scene Would it not be unwise and perhaps even irresponsible in this situation to decide whether to keep this baby-sitter or not by completely ignoring the fact that the police usually run DNA searches trough its DNA database that contains samples from a disproportionately high number of potentially dangerous individuals?

II THE INTERROGATOR'S FALLACY

The name "interrogator's fallacy" was introduced into the literature by Robert A J Matthews[9] He defines this fallacy as "the error of assuming that confessional evidence can never reduce the probability of guilt " Now it is true that we think that the existence of confession normally increases the probability of guilt And we are right about this But Matthews is also right that under certain circumstances the probabilistic impact of a confession could actually go in the opposite direction and *make the guilt less likely*

The crucial question is when exactly the fact of a confession is an indicator of guilt and when not Matthews gives the following answer confession increases the probability of guilt *only* when the probability of confessing given that the person is guilty is higher than the probability of confessing given that the person is innocent If G refers to the statement that a person is guilty and C to the statement that the person confesses, then what Matthews is saying is that $p(G/C) > p(G/{\sim}C)$ only under the condition that $p(C/G) > p(C/{\sim}G)$

---

[9] Matthews, "Inference with Legal Evidence Common Sense Is Necessary but Not Sufficient," *Medicine, Science and Law*, XLIV (2004) 189–92, Matthews, "The Interrogator's Fallacy," *Bulletin of the Institute of Mathematics and its Applications*, XXXI (1994) 3–5

The mathematician Ian Stewart defended the same claim in an article "The Interrogator's Fallacy" in *Scientific American* (in the widely read column "Mathematical Recreations" that he inherited from Martin Gardner) Stewart also argues that "the existence of a confession increases the probability of guilt if and only if an innocent person is less likely to confess that a guilty one "[10] Matthews and Stewart have no doubts about the truth of their claim because they think that it is derivable from some elementary probability formulas And indeed, they can rely on the following statement that is known as the odds form of the Bayes's theorem

$$\frac{p(G/C)}{p({\sim}G/C)} = \frac{p(G)}{p({\sim}G)} \times \frac{p(C/G)}{p(C/{\sim}G)} \tag{5}$$

On the left side of (5) is $p(G/C)/p({\sim}G/C)$, the ratio of so-called *posterior* probabilities of guilt and innocence (that is, probabilities *after* the confession is taken into account) It is equal to the product of two other ratios on the right side of (5) The ratio $p(G)/p({\sim}G)$ represents the ratio of *prior* probabilities of guilt and innocence (that is, probabilities of guilt and innocence *before* the fact of confession is factored in) The last ratio, $p(C/G)/p(C/{\sim}G)$, is called the likelihood ratio and obviously it will determine what impact the new evidence (confession) will have on the probabilities of guilt and innocence In particular, the posterior probability of guilt will be higher than the prior probability of guilt only if the likelihood ratio is larger than 1 But this is exactly what Matthews and Stewart are telling us

In addition, they argue that in some situations the likelihood ratio, $p(C/G)/p(C/{\sim}G)$, will be smaller than 1, and that in these cases a confession will actually become an indicator of *innocence* Their example due to their training and fanaticism, hardened terrorists are in a better position to resist rough interrogation than innocent people who, being much more compliant and suggestible, are unable to cope with the unpleasant situation and often end up by confessing to a crime that they did not commit

---

[10] Stewart, "The Interrogator's Fallacy," *Scientific American* (September 1996) 134–36, Stewart, *The Magical Maze Seeing the World through Mathematical Eyes* (New York Wiley, 1997), pp 90–93

I agree, of course, that (5) is a logical consequence of the axioms of probability theory I also agree that in the example with terrorists the likelihood ratio could indeed be lower than 1 But I do not agree that it necessarily follows from all this that the existence of confession in that situation would be evidence against guilt Contrary to what Matthews and Stewart are saying, I think that *even if an innocent person is more likely to confess than a guilty one, the confession could still raise the probability of guilt*

How can this be? Such an idea seems to be a logical impossibility if we look at (5) But my point is that equation (5) may not be the right formula to use here There could be another empirical fact that is relevant for the probability of guilt and that escaped our attention In that case we would obviously have to use a more complicated formula that would incorporate that fact

Recall how in the case of the prosecutor's fallacy we initially focused exclusively on the DNA match and how at first we completely missed the fact that the defendant's DNA match presupposes the existence of the defendant's DNA in police databases, a very relevant thing for estimating the probability of guilt

Well, I think that Matthews and Stewart are again missing the possible (and moreover probable) existence of something very important in this context the very fact of interrogation Why would this be relevant? Simply because there is an obvious possibility that among those interrogated by the police there is a higher proportion of criminals than in the general population

Since we are examining here the evidential impact of confession *that is the result of interrogation,* the probability we are looking for is not really the probability of *X* being guilty given that he confessed, but rather the probability of *X* being guilty given that he confessed *and that he was interrogated* Therefore, the ratio of posterior probabilities should be expressed in the following more comprehensive way (with '*I*' standing for "having been interrogated")

$$(6)\qquad \frac{p(G/C\,\&\,I)}{p(\sim G/C\,\&\,I)} = \frac{p(G/I)}{p(\sim G/I)} \times \frac{p(C/G\,\&\,I)}{p(C/\sim G\,\&\,I)}$$

or even better, by expanding further the first probability ratio on the right side of (6)

$$(7)\qquad \frac{p(G/C\,\&\,I)}{p(\sim G/C\,\&\,I)} = \frac{p(G)}{p(\sim G)} \times \frac{p(I/G)}{p(I/\sim G)} \times \frac{p(C/G\,\&\,I)}{p(C/\sim G\,\&\,I)}$$

To see why the first ratio on the right side of (6) is equivalent to the product of the first two ratios on the right side of (7), consider first the following two elementary truths in probability

$$(8) \qquad p(G/I) = \frac{p(G) \times p(I/G)}{p(I)}$$

$$(9) \qquad p(\sim G/I) = \frac{p(\sim G) \times p(I/\sim G)}{p(I)}$$

Now if (8) is divided by (9), the magnitude p(I) in the denominators on the right side cancels out and we obtain

$$(10) \qquad \frac{p(G/I)}{p(\sim G/I)} = \frac{p(G)}{p(\sim G)} \times \frac{p(I/G)}{p(I/\sim G)}$$

Equation (10) proves that (6) and (7) are indeed equivalent.

Let us compare equation (5), used by Matthews and Stewart, with equation (7) that gives a more complete picture of the situation. On the left side in each case is the ratio of *posterior* probabilities of guilt and innocence, with the difference that in (5) these probabilities are conditioned only on confession, while in (7) they are conditioned on both confession and interrogation. The first ratio on the right side is the same in both (5) and (7) the ratio of prior probabilities of G and $\sim G$. The last ratio on the right side is also similar in the two cases the probability of confession given guilt (and interrogation) divided by the probability of confession given innocence (and interrogation).

There is a completely new element in (7), however, and it is the middle ratio on the right side $p(I/G)/p(I/\sim G)$. In words, this is the probability of being interrogated by police given that the person is guilty divided by the probability of being interrogated by police given that the person is innocent. We have good reasons to believe that this ratio is higher than 1, unless the police are very incompetent. For it stands to reason to expect that the police will more probably select for interrogation someone who is guilty than someone who is innocent. Furthermore, it seems reasonable to assume that the rougher the interrogation methods used by the police, the more important it will be for them to make sure that those selected for that treatment are indeed guilty.

It is precisely this new element in (7) that reveals what is wrong with Matthews's and Stewart's claim that the probability of guilt can increase after confession *only if* $p(C/G)$ is higher than $p(C/\sim G)$.

Here is a counterexample to their claim. Let us accept their assumption that hardened criminals less frequently confess under police interrogation than innocent people. For the sake of concreteness, let us suppose that, say, only 40 percent of guilty people confess while 60 percent of innocents do Since this makes the probability of confession given guilt *lower* than the probability of confession given innocence, it would seem to follow (as we are told by Matthews and Stewart) that now the existence of a confession cannot increase the probability of guilt from what it was before the confession But suppose further that the police are 9 times more likely to interrogate a guilty person than an innocent individual Finally, suppose that the prior probability of a particular person (let us call him John) being guilty, based on all the evidence before the confession, is 0 75

Here are the probabilities that apply to the situation as described

$p(C/G\&I) = 0.4$

$p(C/{\sim}G\&I) = 0.6$

$p(G) = 0.75$

$p({\sim}G) = 0.25$

$p(I/G)/p(I/{\sim}G) = 9$

Substituting these values into equation (7) yields

$$\text{(11)}\quad \frac{p(G/C\,\&\,I)}{p({\sim}G/C\,\&\,I)} = \frac{p(G)}{p({\sim}G)} \times \frac{p(I/G)}{p(I/{\sim}G)} \times \frac{p(C/G\,\&\,I)}{p(C/{\sim}G\,\&\,I)} = \frac{0.75}{0.25} \times 9 \times \frac{0.4}{0.6} = 3 \times 9 \times \frac{2}{3} = 18$$

Equation (11) tells us that the probability of John being guilty, *all things considered,* is 18 times higher that the probability that he is innocent Since he must be either guilty or innocent, we can use this ratio to obtain the value for the posterior probability of his guilt The probability that he is guilty given that he confessed (and that he was interrogated) is 18/19, or 0 95

Before confession, the probability of John's guilt was 0 75 After confession, the probability of his guilt increases to 0 95 *despite the fact that the probability that he confesses if he is guilty is lower than the probability that he confesses if he is innocent* This completes the proof that Matthews and Stewart are wrong when they insist that confession can be evidence of guilt *only if* the probability of confession given guilt is higher than the probability of confession given innocence

III CONCLUSION

We have seen that some experts in mathematical statistics have given seriously misguided advice about how to avoid fallacious reasoning about probabilities They try to cure us from the prosecutor's fallacy and the interrogator's fallacy but it turns out that their medicine is sometimes worse than the disease

Yet as grave as their lapses of judgment may have been, I surely cannot on that ground alone endorse W H Auden's stern admonition "Thou shalt not sit with statisticians " But a milder version of that warning might be in order If you really must sit with statisticians, do not let them always persuade you too quickly and do not give up your common sense opinion as soon as it clashes with their analysis For as demonstrated in this article, on some rare occasions your gut intuition about probabilities may happen to be more trustworthy than what these professors of statistics are trying to teach you

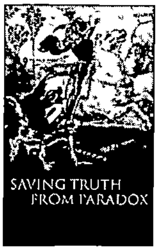
CARMEN DE MACEDO

Hong Kong

New from OXFORD

Saving Truth From Paradox

HARTRY FIELD



This is an ambitious investigation into paradoxes of truth and related issues, with occasional forays into notions such as vagueness, the nature of validity, and the Godel incompleteness theorems Hartry Field presents a new approach to the paradoxes and provides a systematic and detailed account of the main competing approaches

2008 384 pp
cloth $110 00 paper $40 00

The Meaning of Life

A Very Short Introduction

TERRY EAGLETON

"The name Terry Eagleton assures us of stimulation, style, sparkling, sometimes acerbic, wit, and wide-ranging erudition In other words he is eminently readable A commendably pocket-sized book"–Gordon Parsons, *Morning Star*

(Very Short Introductions)

2008 144 pp , 12 halftones paper

Contextuality in Practical Reason

A W PRICE

A W Price explores the varying ways in which context is relevant to our reasoning about what to do He investigates the role of context in practical inferences, practical judgments, and the ascription of reasons for action This original, wide-ranging book will reward the attention of philosophers working in ethical theory and related areas

2008 256 pp $70 00

NEW IN PAPERBACK

Propositions, Functions, and Analysis

Selected Essays on Russell's Philosophy

PETER HYLTON

The prize-winning Russell scholar Peter Hylton presents some of Russell's most celebrated essays from the last two decades, all of which strive to recapture and articulate Russell's monumental vision Relating his work to that of other philosophers, particularly Frege and Wittgenstein, and featuring a previously unpublished essay and a helpful new introduction, the volume will be essential for anyone engaged with the history of twentieth-century ideas

2005 (paper 2008) 230 pp
paper $35 00 cloth $65 00

NEW IN PAPERBACK

Reading Seneca

Stoic Philosophy at Rome

BRAD INWOOD

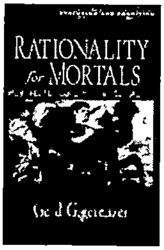
Brad Inwood presents a selection of his most influential essays on the philosophy of Seneca, the Roman Stoic thinker, statesman, and tragedian of the first century AD Including two brand-new pieces, and a helpful introduction to orient the reader, this volume will be an essential guide for anyone seeking to understand Seneca's fertile, wide-ranging thought and its impact on subsequent generations

2005 (paper 2008) 300 pp
paper $38 99 cloth $125 00

Rationality for Mortals

How People Cope with Uncertainty

GERD GIGERENZER



This volume collects Gigerenzer's recent articles on the psychology of rationality This volume should appeal to a broad mixture of cognitive psychologists, philosophers, economists, and others who study decision making

(Evolution and Cognition Series)

2008 256 pp , 23 halftones $65 00

Prices are subject to change and apply only in the US To order, or for more information, please call 1-800-451-7556 In Canada, call 1-800-387 8020 **Visit our website at www oup com/us**

OXFORD
UNIVERSITY PRESS

THE JOURNAL OF PHILOSOPHY

Special Issues Available

Parts and Wholes — December 2006
On Bernard Williams — June 2003
Causation — April 2000
Jurgen Habermas and John Rawls — March 1995

The Dewey Lectures
Michael Dummett — *Truth and the Past*, 2002
Hilary Putnam — *Sense, Nonsense, and the Senses An Inquiry into the Powers of the Human Mind*, 1994
Donald Davidson — *The Structure and Nature of Truth*, 1990
Amartya Sen — *Well-being, Agency, and Freedom*, 1985
John Rawls — *Kantian Constructivism in Moral Theory*, 1980
Ernest Nagel — *Teleology Revisited*, 1977
W V Quine — *Ontological Relativity*, 1968

The Woodbridge Lectures
John McDowell - *Having the World in View Sellars, Kant, and Intentionality*, 1997

Subscriptions

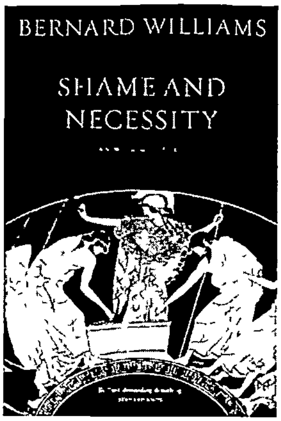
Individuals $45 00
Students $20 00
Libraries and Institutions $100 00
Print and Online for Libraries and Institutions $180 00
Postage outside the United States $15 00

Special Offer

One complimentary issue of any of the above issues with each new subscription
All issues are $5 00 plus $3 00 for shipping and handling
Send check or money order to

The Journal of Philosophy
Columbia University
1150 Amsterdam Avenue, Mail Code 4972
New York, NY 10027

www.journalofphilosophy org



Shame and Necessity

BERNARD WILLIAMS

With a New Foreword by A A Long

"A dazzlingly clever and agile assault Williams's treatment of shame is brilliant Essentially, it is a model of philosophical lucidity And though it is deeply serious, we can often catch an ironic inflection in the author's voice "

—*New York Times Book Review*

Sather Classical Lectures, $21 95 paper

Personal Identity

JOHN PERRY, EDITOR

Second Edition

This volume brings together the vital contributions of distinguished past and contemporary philosophers to the important topic of personal identity New to the second edition are Shoemaker's seminal essay "Persons and Their Pasts," selections from the important and previously unpublished Clark-Collins correspondence

Topics in Philosophy, $22 95 paper

This Is Not a Pipe

MICHEL FOUCAULT

Translated and Edited by James Harkness
Illustrations and Letters by René Magritte
25th Anniversary Edition

"This essay not only proposes a new understanding of Magritte, it also constitutes a perfect illustration and introduction to the thought of the philosopher himself, France's great wizard of paradox "

—*New York Times Book Review*

$15 95 paper

Mental Illness and Psychology

MICHEL FOUCAULT

Foreword by Hubert Dreyfus

Mental Illness and Psychology is an important document tracing the intellectual evolution of this influential thinker A foreword by Foucault scholar Hubert Dreyfus situates the book within the framework of Foucault's entire body of work

$16 95 paper

Philosophical Hermeneutics

HANS-GEORG GADAMER

Translated and Edited by David E Linge
30th Anniversary Edition

"This volume presents carefully selected essays from Gadamer's *Kleine Schriften* Linge does an excellent job in his introduction to Gadamer's thought which precedes the essays "—***Library Journal***

$24 95 paper

At bookstores or order (800) 822-6657 • www.ucpress.edu

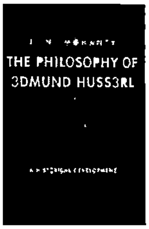
UNIVERSITY OF CALIFORNIA PRESS

Philosophy from YALE

The Philosophy of Edmund Husserl

A Historical Development

J N MOHANTY



"This book crowns Mohanty's fifty-eight years of Husserl studies His scholarship, clarity, and nuanced understanding of the philosophical issues have given us **a great book on a great philosopher."**—Dagfinn Føllesdal

Yale Studies in Hermeneutics series

9 illus $55 00

Leibniz on the Trinity and the Incarnation

Reason and Revelation in the Seventeenth Century

MARIA ROSA ANTOGNAZZA

Translated by Gerald Parks

This book, the first to systematically examine Leibniz's scattered writings on the Trinity and Incarnation, is **"an elegant work of philosophical scholarship."**
—Robert Adams

$60 00

The Case for Greatness

Honorable Ambition and Its Critics

ROBERT FAULKNER

"The classic philosophers, Faulkner contends, help us far more than modern theoreticians to make sense of genuinely great men and women The chapter on Aristotle's 'megalopsychos' is **the best I have read in any of the literatures."**—Steven Smith

$30 00

New in paper

The Republic

PLATO

Translated and with an Introduction by R E Allen

A major new translation of Plato's greatest dialogue.

"Students mystified by the old Cornford translation might find this one easier to grasp"
—*Library Journal*

$20 00 **paperback**

Plato's *Republic*

A Study

STANLEY ROSEN

"The most comprehensive and detailed commentary on and interpretation of Plato's *Republic* in English Stanley Rosen is one of the most original interpreters of Plato alive"—Drew A Hyland

$20 00 **paperback**

European Political Thought 1450–1700

Religion, Law and Philosophy

HOWELL LLOYD, GLENN BURGESS and SIMON HODSON

Drawing on the work of fourteen distinguished historians, this **wide-ranging** book explores political thinking in Europe from the Renaissance to the early Enlightenment

$85 00

YALE University Press • yalebooks com

CAMBRIDGE

New and Noteworthy Titles

Leibniz
An Intellectual Biography
Maria Rosa Antognazza
$39.95 Hb 978 0-521-80619 0
644 pp

The Cambridge Companion to Hegel and Nineteenth-Century Philosophy
Edited by
Frederick C. Beiser
Cambridge Companions to Philosophy
$90.00 Hb 978 0 521 83167 3
$32.99 Pb 978 0-521-53938 8
472 pp

The Cambridge Companion to Hume 2nd Edition
Edited by
David Fate Norton and Jacqueline Taylor
Cambridge Companions to Philosophy
$95.00 Hb 978-0 521-85986 8
$32.99 Pb 978 0 521 67734-9
400 pp

The Cambridge Companion to the "Origin of Species"
Edited by
Michael Ruse and Robert J. Richards
Cambridge Companions to Philosophy
$90.00 Hb 978 0 521-87079 5
$28.99 Pb 978 0-521-69129 1
432 pp

The Philosophy of Simone de Beauvoir
Ambiguity, Conversion, Resistance
Penelope Deutscher
Ideas in Context
$90.00 Hb 978 0 521 88520 1
224 pp

Fellow-Feeling and the Moral Life
Joseph Duke Filonowicz
$99.00 Hb 978 0 521-88871-4
272 pp

The Morality of Embryo Use
Louis M. Guenin
$90.00 Hb 978 0 521-87269 0
$29.99 Pb 978 0 521 69427 8
288 pp

Nietzsche's 'On the Genealogy of Morality'
An Introduction
Lawrence J. Hatab
Cambridge Introductions to Key Philosophical Texts
$90.00 Hb 978 0-521-87502-8
$29.99 Pb 978 0 521-69770 5
296 pp

The Human Right to a Green Future
Environmental Rights and Intergenerational Justice
Richard D. Hiskes
$90.00 Hb 978 0-521 87395-6
$29.99 Pb 978 0 521 69614 2
200 pp

Kierkegaard and Socrates
A Study in Philosophy and Faith
Jacob Howland
$27.99 Pb 978 0 521-73036 5
248 pp

Causation in International Relations
Reclaiming Causal Analysis
Milja Kurki
Cambridge Studies in International Relations
$99.00 Hb 978 0-521-88297-2
$34.99 Pb 978 0 521 70950-7
368 pp

Naturalized Bioethics
Toward Responsible Knowing and Practice
Edited by
Hilde Lindemann, Marian Verkerk, and Margaret Urban Walker
$85.00 Hb 978 0-521-89524-8
$27.99 Pb 978 0 521-71940 7
300 pp

The Mind of Jihad
Laurent Murawiec
$80.00 Hb 978 0 521 88393-1
$24.99 Pb 978 0 521-73063-1
352 pp

Hegel's Practical Philosophy
Rational Agency as Ethical Life
Robert B. Pippin
$90.00 Hb 978-0-521-42957-3
$29.99 Pb 978 0 521-72872-0
312 pp

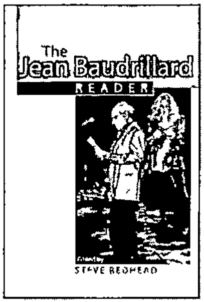
Philosophy in a New Century
Selected Essays
John R. Searle
$90.00 Hb 978-0-521 51591-7
$29.99 Pb 978 0-521-73158-4
220 pp

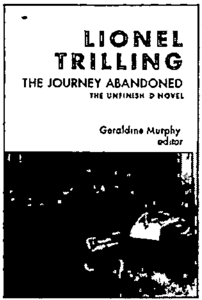
Living with Uncertainty
The Moral Significance of Ignorance
Michael J. Zimmerman
Cambridge Studies in Philosophy
$90.00 Hb 978-0-521-89491-3
240 pp

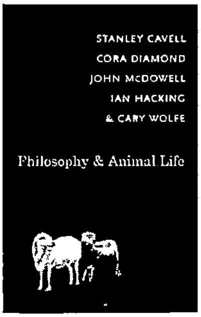
Prices subject to change

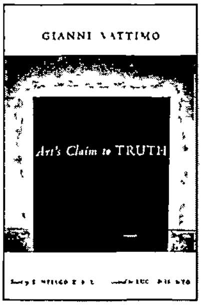
www.cambridge.org/us/philosophy

CAMBRIDGE
UNIVERSITY PRESS









COLUMBIA

Read book excerpts at www.cup.columbia.edu

The Jean Baudrillard Reader
Edited by
Steve Redhead
paper $24.50 / cloth $79.50
European Perspectives
A Series in Social Thought
and Cultural Criticism

Philosophy and Animal Life
Stanley Cavell,
Cora Diamond,
John McDowell,
Ian Hacking,
& Cary Wolfe
cloth $24.50

Animals as Persons
Essays on the Abolition
of Animal Exploitation
Gary L. Francione
cloth - $40.00

Zoographies
The Question of the Animal
from Heidegger to Derrida
Matthew Calarco
paper $24.50 / cloth - $74.50

Democracy and the Political Unconscious
Noelle McAfee
cloth $45.00
New Directions in Critical Theory

The Force of the Example
Explorations in the
Paradigm of Judgment
Alessandro Ferrara
cloth $40.00
New Directions in Critical Theory

The Lives of Sri Aurobindo
Peter Heehs
cloth $45.00

The Journey Abandoned
The Unfinished Novel
Lionel Trilling
Edited by Geraldine Murphy
cloth $26.95

Art's Claim to Truth
Gianni Vattimo
Edited by Santiago Zabala
Translated by Luca D'Isanto
cloth $29.50
Columbia Themes in Philosophy,
Social Criticism, and the Arts

Music, Madness, and the Unworking of Language
John T. Hamilton
cloth - $40.00
Columbia Themes in Philosophy,
Social Criticism, and the Arts

The Hermeneutic Nature of Analytic Philosophy
A Study of Ernst Tugendhat
Santiago Zabala
Foreword by Gianni Vattimo
Translated by Santiago Zabala
with Michael Haskell
cloth $37.50

NOW IN PAPER

The Rise and Fall of Soul and Self
An Intellectual History of
Personal Identity
Raymond Martin &
John Barresi
paper $22.95

Insurmountable Simplicities
Thirty-nine Philosophical
Conundrums
Roberto Casati &
Achille Varzi
paper - $26.00